DISRUPTION OF POSTNATAL NEURONAL SPECIFIC GENE AK045681 INTERFERES WITH ANXIETY BEHAVIOR IN MICE

by

Hideaki Tomita

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Neuroscience Program

The University of Utah

December 2011

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Hideaki Tomita**

has been approved by the following supervisory committee members:

Mario R. Capecchi	, Chair	**5/13/2011** Date Approved
Erik M. Jorgensen	, Member	**5/13/2011** Date Approved
Karl Gordon Lark	, Member	**5/13/2011** Date Approved
Thomas N. Parks	, Member	**5/13/2011** Date Approved
Monica L. Vetter	, Member	**5/13/2011** Date Approved

and by **Mary T. Lucero** , Chair of the Department of **Neuroscience**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Molecular and genetic technology has been essential to answer critical questions in molecular biology. Although many important questions must be answered in molecular biology, advanced molecular tools are yet to be available to answer those critical questions. This thesis describes the development of new genetic tools and their applications. First, the attempt of developing a new genetic technology termed chromosome translocation technology is described. The chromosome translocation technology is a genetic tool that allows us to transfer a large fragment of chromosome between species. A canine quantitative trait locus (QTL) that determines the length and width of femur is used as my model case and transfers this canine QTL to mouse embryonic stem (ES) cells to recapitulate the skeletal phenotype observed in dogs and further characterize this locus in mice. Later, application of *piggyback* DNA transposon system as a mutagenesis tool and characterization of a novel gene AK045681 in mouse from the mutagenesis screen are described. Amino acid sequence similarity of AK045681 to the spliceosomal protein U1C suggested the biological function of this novel gene might be RNA splicing. Double immunostaining showed the co-localization of AK045681 to a core component of the spliceosome SmB/B'/N, which indicated that AK045681 might be a part of the spliceosome. Expression of this novel gene began postnatally and continued in aulthood. AK045681 was mainly expressed in GABAergic neurons in the neural regions that were known to regulate fear or anxiety including the

hippocampus, the amygdala, the basal ganglia and the cortex. Behavioral analyses further revealed that AK045681 knockout mice showed reduced fear or anxiety in the elevated plus maze and dark-light transition task. The results suggested that AK045681 played a role in the regulation of anxiety (fear) in mice. AK045681 was now named *Sango* (meaning after birth in Japanese) after the postnatal-specific expression of the gene.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Molecular and genetic technology has been essential to answer critical questions in biology. Though many important biological questions must be answered, the necessary technological advances are yet limited and the new invention of molecular technologies are long awaited. A goal of this thesis is to describe the challenges in developing new genetic tools and their application. The first half of my thesis describes the attempt to develop a technology (chromosome translocation) that allows us to transfer a large fragment of chromosome between species. A canine quantitative trait locus (QTL) that determines the length and width of femur is used as my model and transfers this canine QTL to mouse embryonic stem (ES) cells to recapitulate the skeletal phenotype observed in dogs and further characterize this locus in mice. To achieve chromosome translocation between species, gene targeting as well as a novel technology, iPSC technology (**i**nduced **P**luripotent **S**tem **C**ell) pioneered by Sinya Yamanaka were utilized as a part of the strategy (Takahashi & Yamanaka, 2006). DNA transposons like *piggyBac* transposons have been shown to be an effective genetic tool for insertion mutagenesis in invertebrates. Recently, *piggyBac* transposons have shown their high mutagenesis efficiency in vertebrates such as mice. However, detailed phenotypic analyses of the mutant mice from the *piggyBac* mutagenesis have rarely been performed.

The second half of my thesis presents successful application of *piggybac* DNA transposon system that has led me to characterize a novel gene locus AK045681 in mice. In this chapter, mechanisms and applications of three genetic tools, gene targeting, iPSC technology and *piggyBac* DNA transposons are reviewed.

Gene targeting and Cre-Lox recombinase system

Gene targeting by homologous recombination in mammalian cells has undoubtedly revolutionized science. The basic principle of gene targeting is that the incoming DNA includes a portion with the same nucleotide sequence as a portion in the targeting gene, therefore enabling homologous recombination between them (Lewin, 2004). Significantly, gene targeting offers the specificity for introducing mutation in mice. With the sequenced genomes such as mice and humans, a specific gene of interest can be altered in any way desired. Additionally, gene targeting allows researchers to gain full control of how to manipulate the gene of interest, and thus give the researchers an opportunity to devise the best genetic modification (e.g., generation of null mutation, reporter allele for cell lineage analysis or tissue-specific Cre-expressing allele for tissue-specific gene inactivation) of the chosen gene to answer a specific biological question (Capecchi, 2001, 2005; Koller & Smithies, 1992).

Gene targeting in mice was pioneered independently by Mario Capecchi and Oliver Smithies following discovery of the mouse embryonic stem cell lines by Martin Evans (Koller & Smithies, 1992). Hereupon, I focus on the Capecchi's work to describe the development of gene targeting. While he was developing a microinjection method as a DNA delivery method into the nuclei of mammalian cells, he observed the multiple

copies of injected DNA sequences were integrated as a highly ordered head to tail concatemer into nuclei of the mammalian cultured cells upon an injection of multiple copies of DNA plasmid (i.e., HSV-tk plasmids) into cells (Capecchi, 2001, 2005). He hypothesized that the highly ordered concatemers should be created in highly regulated fashions such as homologous recombination in cells, and later his laboratory demonstrated that homologous recombination indeed generated those head-to-tail concatemers in the nuclei of mouse fibroblasts and mammalian cells were equipped with an efficient enzymatic machinery to perform homologous recombination between newly introduced exogenous DNAs (Folger et al., 1982; Capecchi, 2001, 2005).

Following this major success, Capecchi and colleagues ultimately deactivate or "knockout" the endogenous hypoxanthine phosphoribosyl transferase (Hprt1) gene by replacing exon 8 of Hprt1 with Neomycin resistance gene (Neo^r) through homologous recombination in mouse ES cells (Thomas and Capecchi, 1987). Their targeting strategy of Hprt1 was ingenious for two reasons. First, Hprt1 gene is located in the X-chromosome, and their mouse ES cell lines were isolated from a male mouse. Thus, gene targeting could easily generate Hprt1 null cell lines by simply disrupting a single copy of the Hprt1 gene. Second, his group exploited the drug selection scheme by 6-thioguanine (6TG), which kills cells with a functional Hprt1 locus (Szybalski, 1992). This is the prelude for invention of the enrichment scheme of gene targeting known as the positive-negative selection, which is Cappechi's other significant contribution to gene targeting technology. His laboratory ultimately perfected the positive-negative selection scheme. In his design he engineered Neo^r gene into his targeting vector, and linked the herpes virus thymidine kinase gene (HVS-tk) to one or both of his targeting vector as positive

and negative selection, respectively. The Neo^r gene was used to ”positively” select for the recipient cells that have integrated the targeting vector into the genome whereas the drug selection by FIAU or ganciclovir, which kills the cells containing a functional HSV-tk gene, selected against (negatively) the cells that have randomly incorporated the targeting vector. Thus, this positive-negative selection scheme dramatically improved a chance to obtain ES cell lines with a properly targeted locus of their choice. Reportedly, enrichment of properly targeted ES cells after double selection with G418 and ganciclovir was two thousand times greater than after single selection of G418 (Mansour et al., 1988). To identify correctly targeted cells among the transfected cells requires researchers to screen as many as 10,000 transfected cells. Thus, the enrichment scheme by positive-negative selection has been extremely beneficial and now is routinely utilized (Koller & Smithies, 1992). Recently, new negative selectable genes such as diphtheria toxin A (DT-A) fragment (DT-A) have been used for a negative selection and shown to be as effective as HSV-tk gene (Araki et al., 2006).

Undeniably, gene targeting by homologous recombination in mice has significantly improved our knowledge of mammalian gene function. However, even this technology encounters a number of limitations because genetic modifications introduced by homologous recombination are permanent in the germ line, and thus those genetic modifications stay in all cells of mice throughout life. Frequently, a number of genes have multiple roles during embryogenesis and adulthood. Knocking out those genes can give rise to an embryonic lethal phenotype, which indicates the earliest nonredundant role of the gene, and prevents an analysis of the gene’s function(s) at later stages of development. Some of the phenotypes caused by such a genetic manipulation are

difficult to interpret because of cell-autonomous, cell-nonautonomous or systemic effects in the whole animal. Generally, permanent germ line mutations by this technology preclude the analysis of gene function in a specific cell type and at given time. In addition, gene targeting by homologous recombination is not apt to engineer complex chromosomal alterations, such as large deletions, duplications, inversions and translocations) that are often linked to human diseases, and thus is not an optimal technique to replicate chromosomal rearrangements that cause human diseases in postnatal life as a result of the interaction of somatic mutation and environmental factors such as sporadic cancer (Ryding et al., 2001; Lewandoski, 2002; Kwan, 2002; Feil, 2007). To circumvent these limitations, a site-specific recombination technology has been built on classic gene targeting by homologous recombination technology.

Site-specific recombination is described by reciprocal exchange between two specific DNA recognition sites regulated by a site-specific recombinase, which creates integration, excision and inversion of defined DNA fragments. The site-specific recombination system generally consists of a pair of DNA recombination sites (approximately 20-200-bps in length) and a specialized site-specific recombinase that recognizes, aligns, breaks and rejoins these sites in reciprocal manner (Ryding et al., 2001;Lewandoski, 2002; Kwan, 2002; Feil, 2007). The recombination sequences are in part asymmetric, conferring directionality to the recombination process. As a result, the outcome of site-specific recombination relies upon the location and relative orientation of the recognition sites regarding one another. If the two sites are on the same DNA molecule, recombination between sites that are the reverse direction triggers an inversion of the DNA between those two sites; as recombination between the two sites that are in

the same direction leads to an excision of the flanked DNA in the form of a circular product (Figure 1-1). If the two sites are on separate DNA molecules, the recombination is intermolecular and can generate DNA integration (translocation) (Figure 1-1). The recombination reaction progress through covalent recombinase-DNA intermediates with conservation of phosphodiester bond energy. As DNA topoisomerases functions, site-specific recombinases break DNA strands by direct phosphoryl transfer to the nucleophilic hydroxyl group of a catalytic tyrosine or serine residue. The cleaved DNA strands are joined together to new partners by reversing the process. In consequence, site-specific recombinses pose dual functions of site-specific endonuclease and ligase. Site-specific recombinases perform the recombination reaction without co-factors as ATP. (Ryding et al., 2001;Lewandoski, 2002; Kwan, 2002; Feil, 2007).

Cre (**c**auses **re**combination) recombinase is 38-kDa protein encoded by bacteriophages P1. Cre recombinase directs recombination between two 34-bp target recombination sites called loxP (**lo**cus of crossing-over [**X**] of **P**1) without any co-factor (Ryding et al., 2001;Lewandoski, 2002; Kwan, 2002; Feil, 2007). The loxP sequence is composed of two 13-bp inverted repeats separated by an 8-bp asymmetric core region that confers overall directionality (Figure 1-1). One Cre monomer binds to each of the palindromes, which facilitates the formation of a synaptic complex of two loxP sites. After the cleavage of DNA strands, four Cre molecules exchange and ligate within the core regions (Figure 1-2). It has been demonstrated that this one site-specific recombination system works particularly well in mice (Ryding et al., 2001;Lewandoski, 2002; Kwan, 2002; Feil, 2007). A classic gene targeting technique is used to generate a mouse in which an essential region of a gene of interest is floxed (flanked by loxP) so

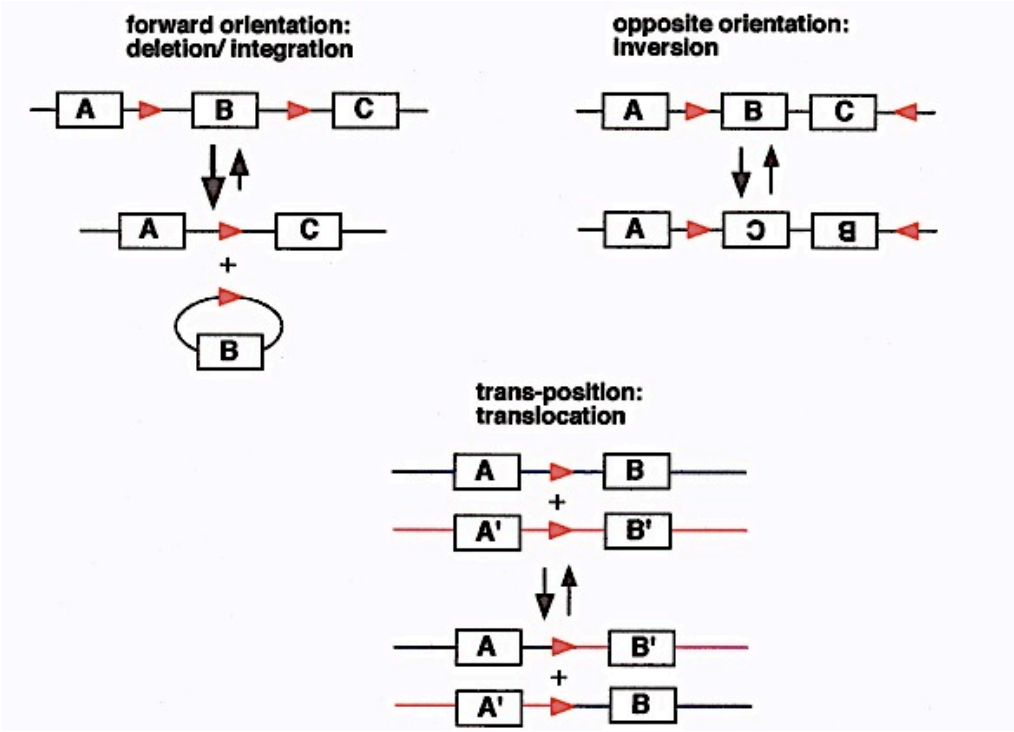

Figure 1-1. Site-specific directional recombination reactions. Site-specific recombinases, such as Cre recobinase, excise a region flanked by the palindromes inter- and intramolecularly, which creates deletion, inversion and translocation (Reprinted with permission from Kwan, 2002).

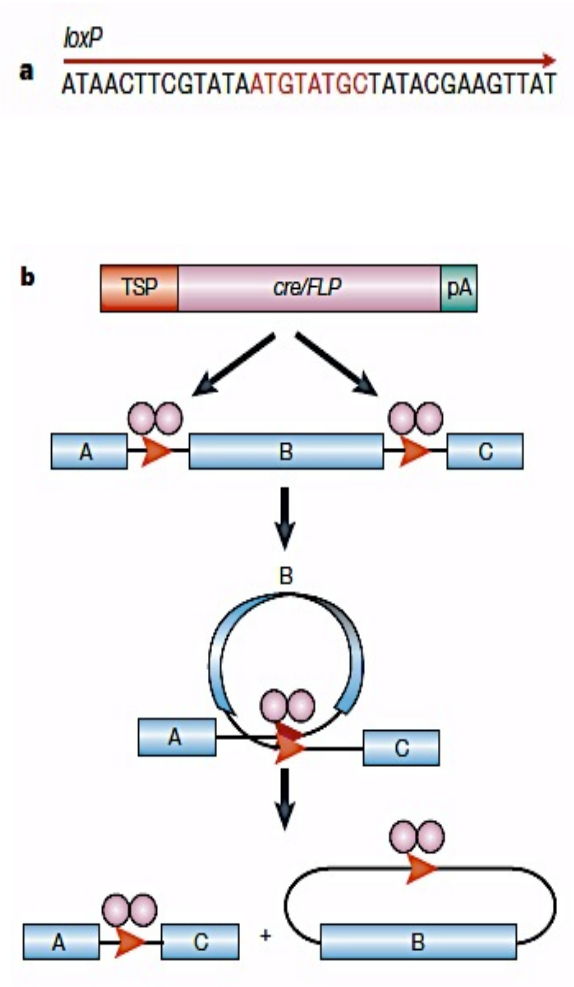

Figure 1- 2. Structure and DNA sequence of the recombinase recognition site loxP (Top Panel). The loxP site is composed of two 13-bp palindromes flanking an asymmetric 8-bp core, which define the orientation of the recognition site. General strategy of Cre-mediated tissue-specific inactivation in mice (Bottom Panel) (Reprinted with permission from Lewandoski, 2001).

that crossing with other mice transgenic for Cre recombinase under the control of a specific promoter allows time- and tissue-specific deactivation of the floxed region of the gene of interest.

This "conditional" knockout/inactivation of genes ensures that the activity of the gene is modified in a limited range of cells at a particular developmental stage, whereas the cells in the rest of the animal's body are genetically unaltered. Lakso and colleagues (1992) first demonstrated that the Cre-regulated activation of SV40 large T antigen transgene in lens cell led to tumor formation (Figure 1-2 for general description of tissue-sepcific KO). Furthermore, Rajewsky and colleagues further improved the Cre/lox-mediated tissue-specific recombination (Gu et al., 1994) and inducible recombination in mice (Kuhn et al., 1995). Numerous conditional mouse lines have been made to elucidate tissue-specific functions of genes (see Kwan, 2002 for the list of conditional mice lines available).

Cre/lox strategy can be effectively used for cell fate-mapping studies. Ce/lox-based cell fate mapping uses reporter mice in which a histological marker can be activated by Cre-mediated recombination in nearly any tissue, thus reporting Cre activity. The generation of such reporter mouse lines can be accomplished by classic gene targeting to Cre-activatable reporter cassette (β-galactosidase, enhanced fluorescent protein of green, yellow or cyan varieties) into Rosa26 locus, which generates constitutive transgene expression. Those reporter mice lines are effective for linking gene expression to the fate of particular cell lineage because activation of the reporter gene by a specific tissue-specific Cre-expressing mouse line suggests that Cre activity is present in that particular cell or was present in a progenitor cell (Ryding et al., 2001;

Lewandoski, 2002; Kwan, 2002; Feil, 2007). This approach has been successfully used to fate map the mid-hindbrain border (Zinyk et al., 1998), neural-crest cell diversity (Epstein et al., 2000; Chai et al., 2000; Jiang et al., 2000), pancreatic islet cells (Herrera et al., 2000) and memory T cells (Jacob et al., 1999).

To control not only the site of recombination but also the timing, an inducible form of Cre recombines has been developed. The ligand-binding domain of a mutated estrogen receptor is fused to Cre recombinase (Cre-ER^T). This hybrid Cre is only activated by in the presence of tamoxifen but not endogenous steroids (Metzger et al., 1995). Similar hybrid Cre recombinases have been engineered using the ligand-binding domains of a mutated progesterone receptor and glucocorticoid receptor, which respond to RU486 and dexamethasone, respectively (Kellendonk et al., 1996; Brocard et al., 1998). The binding of the cognate ligand (e.g., tamoxifen) to the ligand-binding domains of a mutated steroid receptor (e.g., estrogen) allows translocation of the hybrid Cre (e.g., Cre-ER^T recombinase) into the nucleus in which the hybrid Cre performs the recombination of its floxed DNA substrate, whereas the hybrid Cre stays in the cytoplasm in the absence of ligands. It has been reported that the recombination efficiency was almost 100% in Cre-expressing cells after 3-days exposure of tamoxifen (Brocard et al., 1997) and no detectable recombination was found in the absence of tamoxifen in the Cre-ER^T transgenic mice (Feil et al., 1996). However, some reports have suggested that incomplete and ligand-independent recombination have been identified in some of those ligand-activated Cre transgenic mice (Schwenk et al., 1998; Kellendonk et al., 1999). It has been shown that transgenic mice expressing the Cre-ER^T recombinase require relatively high doses of tamoxifen to induce recombination, which may cause undesired

side effects. As a consequence, novel tamoxifen-inducible Cre recombinases have been sought out to increase sensitivity and efficiency of inducible recombination in mice. Cre-ER^{T2} has been proven to be approximately 10-fold more sensitive to tamoxifen than Cre-ER^{T} (Feil et al., 1997). AT present, the Cre-ER^{T2} recombinase is suggested to be the best tool for the temporal-regulated somatic mutagenesis in mice.

As reviewed above, gene targeting by homologous recombination is the essential genetic tool in science. This technology has been further improved with Cre/lox recombination system, which allows researchers to inactivate gene of interest temporally and spatially at their will. The reporter mouse lines have broadened up the utility of gene targeting. Collectively, gene targeting with Cre/lox recombinase allows researchers to creatively study gene functions in vivo.

Induced pluripotent stem cell (iPSC) technology

Induced pluripotent stem cells (iPSCs) are genetically reprogrammed somatic cells re-acquiring embryonic stem cell-like properties by forced expression of defined factors that are important for the maintenance of stem cells. Yamanaka and colleagues (2006) designed an elegant screen for "reprogramming" factors out of 24 stem cell-related candidate genes that could activate the genetically engineered dormant drug resistance allele in the embryonic stem cell (ESC)-specific *Fbxo15* locus. Their screen identified the combination of 24 factors that activate Fbxo15 and generate the drug-resistant colonies with distinctive ESC morphology when those 24 factors were expressed from retroviral vector in mouse embryonic fibroblasts (Takahashi & Yamanaka, 2006). Multiple rounds of screenings eventually identified the minimal set of four genes, *Oct4*,

Sox2, *Klf4* and c-*Myc* that are required for the reprogramming. Those drug-resistant (Fbxo15 expressing) ESC-like colonies are named as **i**nduced **p**luripotent **s**tem **c**ells (iPSCs). Those iPSCs further complied with major criteria of pluripotency, such as expression of pluripotent stem cell markers (SSEA-1, and Nanog), teratoma formation upon the subcutaneous injection into immunocompromised mice, and contribution to different tissues during embryogenesis upon blastocyst injection (Hanna et al., 2010; Stadtfeld & Hochedlinger, 2010; Jaenisch & Young, 2008). However, those iPSCs failed to generate chimeras or contribute to the germline because those iPSCs had insufficient levels of several crucial pluripotent genes in comparison with ESCs, and show only partial demethylation of stem cell regulator such as *Oct4* (Takahashi & Yamanaka, 2006). Following the success, Yamanaka and other laboratories improved the selection strategy by utilizing the essential gene loci including *Nanog* and *Oct4* instead of *Fbxo15*, and successfully reproduced the Yamanaka's initial findings. In fact, iPSCs selected by selection of *Nanog* or *Oct4* reactivation showed closer molecular and functional resemblance to ESCs (Maherali et al., 2007; Okita, et al., 2007; Werning et al., 2007).

Forced expression of four "Yamanaka" factors has successfully generated iPSCs from different species including humans, mice, rats, pig, and rhesus monkeys, which indicates that essential aspects of the transcriptional system regulating pluripotency remains conserved during evolution (Takahashi & Yamanaka, 2006; Takahashi et al., 2007; Liao et al., 2009; Esteban et al., 2009; Liu et al., 2008). Also, the Yamanka factors have been applied to generate iPSCs in a vareity of somatic cell types such as fibroblasts (Takahashi & Yamanaka, 2006), keratinocytes (Maherali et al., 2008), neural cells

(Eminli et al., 2008), and melanocytes (Utikal et al., 2009), as well as genetically labeled pancreatic β cells (Stadtfeld et al., 2008a) and terminally differentiated lymphocytes (Hanna et al., 2008: Eminli, et al., 2009), which further suggests that conservation of transcriptional network of pluripotency.

iPSC technology shows great potentials to be translated into clinical applications such patient-specific cell transplantation therapy. To translate this technology into clinical applications, safe and efficient factor delivery methods must be engineered. Constitutively active retro- or lentiviral vectors are used to introduce Yamanaka factors into somatic cells. Both retro- and lentiviral vectors stably integrate its transgenes into the host genome upon the infection. The retroviral delivery of Yamanaka factors often generates incomplete or partially reprogrammed cells lines that continuously depend on exogenous Yamanaka factor expression and fail to activate the corresponding endogenous factors because the retroviruses are typically shut down toward the end of the reprogramming process as a result of the activation of DNA and histone methyltransferases (Takahashi & Yamanaka, 2006). Moreover, reactivation of the integrated retroviral transgenes in iPSC-derived somatic cells can disrupt the developmental potentials and often result in the formation of tumors in chimeric animals (Okita, et al., 2007). As a delivery system, lentiviral vectors have advantage over retroviral vectors. Lentivirus can infect different somatic cell types more effectively than retrovirus. Also, lentiviral vectors can carry and express all four reprogramming factors as polycistronic cassette, thus increasing reprogramming efficiency. Furthermore, the lentivirus can produce iPSCs more effectively than the retrovirus because the lentiviruses are less efficiently silenced than retroviruses in the infected host cells.

However, at the same time this can cause a differentiation block once the host cells gain pluripotency (Sommer et al., 2009). To overcome a differentiation block in pluripotent cells, constitutively active lentiviral vectors have been further evolved into inducible lentiviral vectors. The use of inducible lentiviral vectors, whose expression can be regulated by doxycycline treatments, significantly decreases the chance of continuous expression of exogenous Yamanaka factors. The use of inducible lentiviral vectors also guarantee truly reprogrammed iPSCs because incomplete or partial reprogrammed iPSCS cease the proliferation upon doxycycline withdrawal (Sommer et al., 2009).

As stated above, the viral delivery (either retro- or lentivirus) of Yamanaka factors poses potentially harmful effects such as the viral reactivation of transgenes and the insertion mutagenesis. To circumvent the possible damaging effects, nonviral or integration-free delivery systems have been sought out. Specifically, it is specifically crucial that iPSC technology is applied to clinical settings. Stadtfeld and colleagues (2008b) demonstrated iPSC generation of adult mouse hepatocytes by using nonintegrating adenoviral vectors. Okita and colleagues (2008) also showed that mouse embryonic fibroblasts were reprogrammed by transfection of DNA plasmids. Furthermore, human fibroblasts have been reprogrammed into human iPSCs by using adenoviral vectors (Zhou and Freed, 2009), Sendai virus (Fusaki et al., 2009) as well as polycistronic mini circle vectors (Jia et al., 2010) and self-replicating selectable episomes (Yu et al., 2009). These reports strongly support that transient expression of four Yamanaka factors is sufficient to induce pluripotentcy in somatic cells. Those reports also validate the absence of common integration sites in iPSCs generated with retroviruses or lentiviruses, which suggests that insertion mutagenesis at best plays a

supportive role during iPSC generation. A significant drawback of nonintegrating delivery system is substantially low reprogramming efficiency. Reprogramming efficiency with the nonintegrating vectors is several folds lower (~0.001%) than that with integrating vectors (0.1% ~ 1%) (Stadtfeld & Hochedlinger, 2010).

It appears that nonintegrating vectors cannot sustain the expression of four Yamanaka factors for an adequate amount of time to achieve complete epigenetic remodeling. To overcome the low reprogramming efficiency of nonintegrating vectors, excisable integration-dependent gene delivery vectors have been explored. The integrating gene delivery vectors, such as lentiviruses, are devised with incorporated loxP sites that can be excised from the host genome upon transient expression of Cre-recombinase. This gene delivery strategy can still generate iPSC effectively from different cell types if polycistronic vectors are utilized (Sommer et al., 2010). Inevitably, this strategy leaves short vector sequences, such as loxP site in the host genome after excision. It still remains to be seen if these short vector sequences have an effect on cellular function of the host. Furthermore, integration-free iPSCs can be generated with *piggyBac* DNA transposons, mobile genetic elements that can be inserted and removed from the host genome by transient expression of transposase. *PiggyBac* transposons are an ideal gene delivery system because this vector leaves no vector sequence behind upon the excision. But integration sites in iPSCs are still needed to be characterized before and after transposons are removed (Kaji et al., 2009; Woltjen, et al., 2009). It also remains unknown if the transposase expression can cause nonspecific genomic modification in iPSCs.

iPSCs have been successfully generated without the use of viral or plasmid vectors. Zhou and colleagues (2009) have reported that human and mouse fibroblasts have been successfully reprogrammed by application of purified recombinant proteins of four Yamanaka factors. In addition, successful reprogramming has been accomplished by delivery of four reprogramming factors as whole-cell extracts isolated from either ESCs (Cho et al., 2010) or genetically engineered HEK293 cells (Kim et al, 2009). The delivery of four reprograming factors as purified proteins appears to be the ultimate way to achieve transgene-free iPSC; however reprogramming efficiency of four recombinant proteins is exceptionally low. The recombinant proteins of four reprogramming factors also required the aid of the histone deacetylase (HDAC) inhibitor valproic acid (VPA) for its successful iPSC generation (Zhou et al., 2009). Recently, modified RAN- and miRNA-based reprogramming approaches, which may be more effective and safer, have successfully achieved iPSC generation from human and mouse somatic cells (Warren et al., 2010; Anokye-Danso et al., 2011).

Whether using integrating or nonintegrating methods, overall reprogramming efficiency of iPSC generation is not as high as one can desire. To improve the overall low efficiency of this technology, chemical compounds that can facilitate reprogramming processes have been researched. Several laboratories have identified numerous molecules that can substantially increase reprogramming efficiency in the presence of the four reprogramming factors (refer to Li and Ding, 2010; Feng et al., 2010 for review). Some of those reprogramming-capable molecules can take the place of individual reprogramming factors, *Oct4*, *Sox2*, *Klf4*, and *c-Myc*, which suggests that those reprogramming-capable molecules can be used solely to derive iPSC formation.

However, there is often a dramatic decrease in the number of iPSC colonies when those reprogramming factors are replaced by the small molecules, which indicates that no single chemical compound is able to fully take the place of a transcription factor as its function. It should be noted that those reprogramming-capable compounds are shown to be powerful modulators of DNA and chromatin modifications. Thus, chemical reprogramming may potentially introduce genetic or epigenetic abnormalities into iPSCs.

Reprogramming inefficiency (0.01% ~ 0.1%) and slow kinetics (~2 weeks) have been the major stumbling blocks of iPSC technology, which suggests that there are chief molecular events that may act as "roadblocks" during the reprogramming processes (Stadtfeld et al., 2008c; Brambrink et al., 2008). Molecular and cellular mechanisms underlying iPSC generation must be elucidated for potential use of this technology in clinical settings in future. Two models have been put forward to explain the low efficiency and slow kinetics of reprogramming. The "elite" model describes only a few cells in a somatic cell culture are amenable to reprogramming and thus the efficiency of reprogramming is significantly low. The most palpable candidate cells are somatic stem cells or progenitor cells in adult tissues and explanted cells populations since they are scarce and developmentally closer to pluripotent. On contrary, the "stochastic" model states that all somatic cells are equally susceptible to reprogramming yet have to undergo a series of stochastic epigenetic events to gain pluripotentcy. Ultimately only a few cells may be able to overcome the epigenetic barriers or roadblocks to acquire pluripotency; therefore, the reprogramming efficiency is low. Both elite and stochastic model are supported by experimental evidence. For instance, iPSCs formation can be achieved from the defined somatic cells types such as pancreatic β cells (Stadtfeld et al., 2008a)

and fully differentiated B and T lymphocytes (Hanna et al., 2008; Eminli et al., 2009). Furthermore, when clonal populations of early β cells monocytes expressing reprogramming factors were followed, a vast majority of cell clones eventually generated the daughter cells that formed iPSCs between several weeks to months (Hanna et al., 2009a). It suggests that rare cells in a homogenous cell population can go through stochastic changes that promote their conversion into pluripotent state with multiple rounds of cell proliferation (Hanna et al., 2009a). On the other hand, it has been demonstrated that clonal populations of hematopoietic stem cells and progenitor cells become iPSCs more efficiently (10% ~ 40%) and faster than mature lymphocytes and myeloid cells (0.01% ~ 1%) in spite of the proliferative state of cells at the time of factor expression (Eminli et al., 2009). This suggests that the differentiation status of cells may be an influential factor for iPSC generation. Intuitively thinking, since adult progenitors and stem cells are developmentally immature, they have to go through fewer stochastic epigenetic events to be reprogrammed than fully differentiated cells, which can explain the faster and more efficient iPSC generation from adult progenitors and stem cells. In fact, those developmentally immature cells up-regulated pluripotency markers faster than terminally differentiated cells during the reprogramming process (Eminli et al., 2009). Thus, a true mechanism of reprogramming efficiency and kinetics may be explained by a combination of those two models.

The reprogramming process appears to be a well-ordered sequence of events. This organized sequence of events begins with down-regulation of somatic-specific genes and mesenchymal-to-epithelial transition (MET)-like morphological changes, which is a crucial early phase of reprogramming as activation of Tgf β signaling, inhibition of BMP

signaling or depletion of MET gene such as E-Cadherin cancel reprogramming (Stadtfeld, et al., 2008c; Li et al., 2010; Mikkelsen et al., 2008; Samavarchi-Tehrani, et al., 2010). Following these events, early pluripotentcy markers SSEA-1, alkaline phosphatase and Fbxo15 become expressed (Stadtfeld et al., 2008c; Brambrink et al., 2008). Cells become independent from exogenous reprogramming factors as true pluripotency genes such as Nanog or Oct4 are activated (Maherali et al., 2007; Stadtfeld et al., 2008a). Distinct differences in telomere length, global transcriptional and DNA methylation patterns have been identified between early- and late-passage iPSCs, which suggests that the acquisition of pluripotency may not be finished upon the transition from exogenous to endogenous expression of four reprogramming factors, and may need several rounds of cell divisions (Marion et al., 2009; Chin et al., 2009; Polo et al., 2010). Genome-wide analyses of histone methylation marks (H3K9me2, H3K4me3, H3K27m33) and DNA methylation showed significant differences between fibroblasts, partial-reprogrammed iPSCs and fully reprogrammed iPSCs and they are re-set to ES cell-like patterns (Maherali et al., 2007; Mikkelson et al., 2008; Chin et al., 2009; Doi et al., 2009; Sridharan et al., 2009; Hawkins et al., 2010). Additionally, it has been demonstrated that asymmetric cysteine methylation, which is extensively observed in ES-cells, is restored in fully reprogrammed iPSCs (Lister et al., 2009). The evidence above indicates that the transition to pluripotent state is associated with a major re-setting chromatin landscape.

It has been shown that down-regulation of somatic-specific markers and activation of embryonic markers are observed in only fraction of fibroblasts expressing reprogramming factors (Wernig et al., 208; Stadtfeld et al., 2008a). This suggests that a

vast majority of cells are refractory to reprogramming or become refractory as a result of ectopic expression of reprogramming factors. Reprogramming intermediates selected based on the combination of pluripotency markers described above have an increased chance of generating iPSC colonies, which suggests that those reprogramming intermediates have indeed passed some transcriptional and epigenetic barriers that normally block the induction of pluripotency (Stadtfeld et al., 2008c). Taken together, the arrest of somatic program and the consequential activation of endogenous pluripotent genes may be a roadblock during iPSC generation.

Among key endogenous pluripotent regulators, activation of endogenous Nanog locus by reprogramming factors may be the most critical since it co-binds numerous targets of *Oct4*, *Sox2*, and *Klf4* in ES cells as well as biochemically interacts with *Oct4* and other transcription factors in ES cells (Wang et al., 2006). Since *Nanog* is up-regulated only late during reprogramming, it has been speculated that *Nanog* may be limiting for efficient transition of somatic cells to iPSCs (Hanna et al., 2010). In agreement with this theory, *Nanog* positively influences reprogramming by cell fusion (Silva et al., 2006). Likewise, *Tgf β* inhibition, which is a critical process for transition to the pluripotent state, may be regulated through up-regulation of *Nanog* (Ichida et al., 2009). Most essentially, it has been shown that *Nanog* is absolutely required for complete induction of pluripotency only during the final stages of the reprogramming process (Silva et al., 2009). Finally, overexpression of *Nanog* as well as other pluripotency transcription factors such as *Sall4* and *Tbx3* has been shown to boost reprogramming efficiency (Hanna et al., 2009a; Tsubooka et al., 2009; Wang et al., 2006; Liang et al., 2008; Yang et al., 2008).

It has been suggested that re-establishment of ES cell-like chromatin state is vital for the reprogramming process. The promoters of pluripotency genes such as *Nanog* and *Oct4* are known to be silenced securely by DNA methylation in somatic cells (Gidekel & Bergman 2002). The promoter DNA methylation profiles of *Nanog* and *Oct4* have been specifically examined between mouse embryonic fibroblasts (MEFs), reprogramming intermediates, iPSCs and ESCs. The promoters of those pluripotency genes are hypermethylated in MEFs and reprogramming intermediates whereas those promoters are unmethylated in iPSCs and ESCs (Maherali et al., 2007; Mikkelsen et al., 2008). In agreement with the idea that DNA methylation interferes with the binding of transcription factors and subsequent gene activation during reprogramming, the reprogramming efficiency is enhanced with the addition of the DNA demethylating drug, 5-azacytidine as well as depletion of maintenance methyl transferase *Dnmt1*(Mikkelsen et al., 2008). It is not clear how these epigenetic roadblocks are surpassed in the process of iPSC formation. However, it has been speculated that passive or active mechanisms may be in place to reduce DNA methylation in the promoters of pluripotency genes (Hockedlinger & Plath 2009). Collectively, this evidence indicates that chromatin states may also be roadblocks in the reprogramming process.

ES cells possess a self-renewal capability, which grants them immortality. Fibroblasts and other somatic cell types are limited in their proliferative potential and undergo apoptosis, growth arrest, and stress-induced senescence in culture. Therefore, the acquisition of cellular immortality could be another biological barrier to the reprogramming process. In fact, serial passages of fibroblasts are highly correlated with a decrease in reprogramming efficiency (Utikal et al., 2009). It has been reported that

increased proliferation is observed at an early phase of fibroblast reprogramming (Mikkelsen et al., 2008). Up-regulation of genes driving cell cycle progression also seems to be an early event during fibroblast reprogramming (Mikkelsen et al., 2008). However, as stated earlier, cells at the intermediate stages of reprogramming have shown the dependence to exogenous factor expression for continuous growth, which indicates that the acquisition of immortality takes place late during the reprogramming process. In agreement with the hypothesis that acquisition of cellular senescence is a roadblock during fibroblast reprogramming, ectopic expression of four reprogramming factors in *p53* or *lnk4a*/*arf* deficient immortalized fibroblasts enhanced the reprogramming efficiency and kinetics (Banito et al., 2009; Hong et al., 2009; Kawamura et al., 2009; Li et al., 2009; Marion et al., 2009; Utikal et al., 2009).

Besides the role in facilitating the reprogramming described above, *p53* loss may also promote reprogramming efficiency by blocking DAN-damaged induced apoptosis (Marion et al., 2009). Interestingly, however, *p53* loss can evoke different responses in cells expressing four reprogramming factors. In fibroblasts, loss of *p53* contributed to reprogramming by inhibiting senescence and cell death whereas in blood cells with ectopic expression of four factors, *p53* loss contributed to reprogramming by accelerating cell cycle progression (Hanna et al., 2009b). Together, this evidence indicates senescence and apoptosis as potential roadblocks to the reprogramming process.

iPSC technology holds enormous potential to translate into therapeutic tools. Organ transplantation among unrelated individuals involves complications such as the limited availability of matched organs and life-long treatments with immunosuppressive drugs that may cause serious side effects. These problems can be overcome since iPSCs

can be differentiated into the desired cell types that are already genetically matched with the patients. The feasibility of cell therapy with iPSCs to genetic disorders has been tested in mice. It has been demonstrated that iPSCs are capable of rescuing the defects in a mouse model of sickle cell anemia (Hanna et al., 2007). Sickle cell anemia is caused by the single point mutation in the hemoglobin gene, which results in the dysfunction of red blood cells. iPSCs were generated from skin cells isolated from the mouse model, which recapitulated the disease phenotypes in humans (Hanna et al., 2007). Afterwards, the disease-causing mutation was corrected in iPSC by gene targeting, and the repaired iPSCs were further differentiated into the blood-forming progenitors. Those "cured" progenitors were later transplanted into anemic mice in which those progenitors generated normal red blood cells and cured the disease. This approach could be applicable to any human disease for which the disease-causing mutation is known and could be treated by cell transplantation. In fact, the phenotypes of the hemophilia A in mice were corrected by transplantation of heterologous, iPSC-derived endothelial progenitors (Xu et al., 2009).

The study and treatment of degenerative disorders such as type I diabetes, Alzheimer's disease and Parkinson's disease also are difficult, because of the availability of the perturbed tissues and the inability to culture those affected cell types for an extended period of time. iPSCs may be used to generate iPSCs from skin cells of the affected patients and later differentiate them in vitro into the perturbed cell types. Thus, the researchers can re-establish the disease in Petri dishes. This "disease-modeling" approach is useful because iPSC can re-create and study the actual cell types that are compromised and lost in the patients (Saha & Jaenisch 2009; Abeliovich & Doege 2009;

Gubaseeli et al., 2010; Koch et al., 2009; Stadtfeld & Hochedlinger 2010). Furthermore, iPSCs can be an unlimited source for any desired specialized cells because iPSCs grow indefinitely in culture. In the long run, this approach can be used as an in vitro screen system and can identify a novel compound to treat the disease including amyotrophic lateral sclerosis (ALS) and spinal muscular atrophy (SMA) (Dimos et al., 2008; Ebert et al., 2009). This approach also allows researchers to re-cerate and capture the very early stage of degenerative diseases since iPSCs have to undergo the same differentiation process in culture as the patients experience during the disease progression. This is especially critical because cellular specimens are generally isolated from the patients who have progressed further to advanced stages of diseases at the time of diagnosis; thus major events of diseases have taken place and been missed. Ultimately, this technology may provide valuable insights into the timing and causes of degenerative disorders.

As reviewed above, iPSC technology has provided a unique opportunity to study detailed mechanism of cellular reprogramming. Indeed, since the invention of this technology, researchers have gained extensive knowledge of cellular reprogramming and stem cell behaviors. Although there are still key improvements to be made such as the invention of safe and efficient gene delivery methods (i.e., integration-free iPSCs with intact epigenetic integrity), this technology holds great promise to be translated into therapeutic tools in the future.

DNA transposon system

Transposable elements or transposons are distinct mobile elements capable of moving around and replicating within the genome of an organism. Transposons were

first identified by Barbara McClintock in the maze genome, and have been identified basically in all living organisms since the discovery (Lewin 2001). For instance, approximately 45% of the human genome is made out of the transposon-derived elements. Nonetheless, a vast majority of the transposons were inactive millions of years ago (International Human Genome Sequencing Consortium 2001). Transposons could be best described as molecular parasites that propagate themselves by using the hosts' resource. Transposons are noninfectious unlike viruses. The parasitic activities are limited within cells, which have forced transposons to develop harmless strategies toward the host in order to coexist and survive with the host (Hartl et al., 1997). In spite of their parasitic nature, it has been suggested that transposons might be a strong force in gene evolution (Izsvak & Ivics 2003). For instance, approximately 50 human genes are originated from transposable elements. Among those there are genes that are responsible for recombination of immunoglobulin gene in all vertebrates (Izsvak & Ivics 2003).

Transposons are categorized into two main classes: retrotransposons and DNA transposons. Retrotransposons are mobile elements transposing through RNA intermediates such as long interspersed elements (LINEs), short interspersed elements (SINEs), and long terminal repeats (LTR) retrotransposons. Retrotransposons utilize a "copy-and paste" mechanism of transposition. The donor elements are not mobile, but reverse transcription replicates the donor elements (copy), which insert (paste) elsewhere in the genome. DNA transposons are mobile elements that can relocate directly as DNA within the genome by a "cut-and-paste" mechanism, in which the element itself is excised from the donor site and is reinserted elsewhere in the genome (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

DNA transposons show great potential as powerful genetic tools because they are safe nonviral vectors that can transfer a define DNA segment from one location to another (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010). In invertebrates such as flies and worms, endogenous elements, such as P-element and Tc1 element for *Drosophila* and *Caenorhabditis elegans* respectively, have been developed as tools for insertional mutagenesis and germline transgenesis (Cooley et al., 1988; Plasterk 1996). Unfortunately, no active endogenous element was available in vertebrates until recently (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010). A Tc1-like active DNA transposon, *Sleeping Beauty* was reconstructed from salmanoid fish and has been shown to be highly active in human and mouse cells. *Sleeping Beauty* has been tested for germ line mutagenesis, but it has not been as effective as it is in vitro (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010). Following the discovery of *Sleeping Beauty*, a new Tc1-like active DNA transposon, *piggyBac* has been isolated from the cabbage looper moth, Trichoplusia ni. *piggyBac* transposons have been shown to be highly active in mouse cells and have demonstrated a significantly high efficiency for germ-line mutagenesis in mice (Handler 2002; Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

Sleeping Beauty transposon is the best-characterized DNA transposon among the known Tc1-like DNA transposons. The structure and transpositional mechanism of *Sleeping Beauty* have been well-studied. I use *Sleeping Beauty* as an example to describe the structure and mechanism of DNA transposition and discuss critical differences in the

structure and transpositional mechanism between *piggyBac* and *Sleeping Beauty*. DNA transposons of *Sleeping Beauty* and *piggyBac* have a simple gene structure (Figure 1-3). In nature, DNA transposons are made of a single gene encoding the transposase polypeptide, the emzymatic factor of transposition, which is flanked by terminal inverted repeats (IRs) that are binding sites for the transposase. In the laboratory, the transposase gene can be physically separated from the IRs, and replaced with any DNA sequence desired. Transposases can mobilize the transposons *in trans* as long as the IRs are present. Transposases can be supplied as a form of another DNA molecule, fused DNA with your cargo, mRNA or protein (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

The terminal inversed repeats (IRs) of both DNA transposons contain two direction repeats (DRs) (Figure 1-3). The distinctive organization of terminal inverted repeats (IRs) is known as IR/DR, which is evolutionarily conserved among the Tc1-like transposable element and is a strict requirement for the transposition. *Sleeping Beauty* transposons have short 15-20-bp imperfect DRs that contain two transposase binding sites. *piggyBac* transposons have 13-bp perfect inverted terminal repeats and 19-bp subterminal repeats located 31-bp from the 5' IR and 3-bp from 3' IR. All four binding sites are required for the transposition. Specific binding to DRs of *Sleeping Beauty* transposase is mediated by an N-terminal, pair-like DNA-binding domain, which consists of PAI and RED subdomain. The RED subdomain partially overlaps with a nuclear localization signal that is flanked by phosphorylation target sites of casein. The phosphorylation of those sites serves as a checkpoint for the regulation of transposition. The nuclear localization signal gives the transposases easy entry to the host by taking

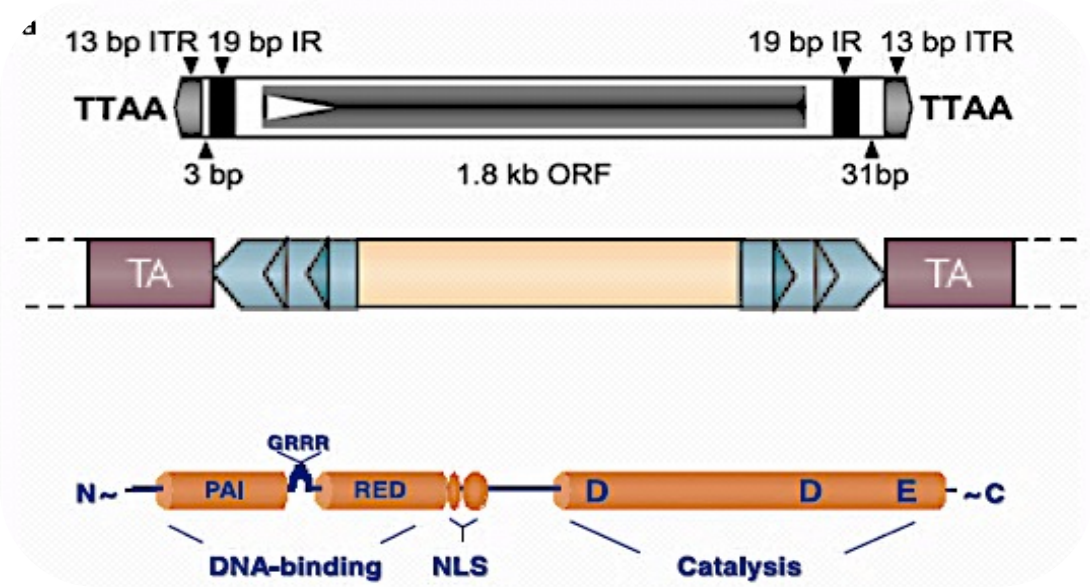

Figure 1- 3. Structure of Sleeping Beauty and PiggyBac transposons (Top Panel). The functional domains of Sleeping Beauty transposable element (Middle Panel). The "cut-and-pasts" transposition mechanism of Sleeping Beauty transposon and a hypothesized functional role of DNA-binding high-mobility group B1 protein (HMGB1) in Sleeping Beauty synaptic complex formation (Bottom Panels)(Reprinted with permission from Copeland & Jenkins 2010).

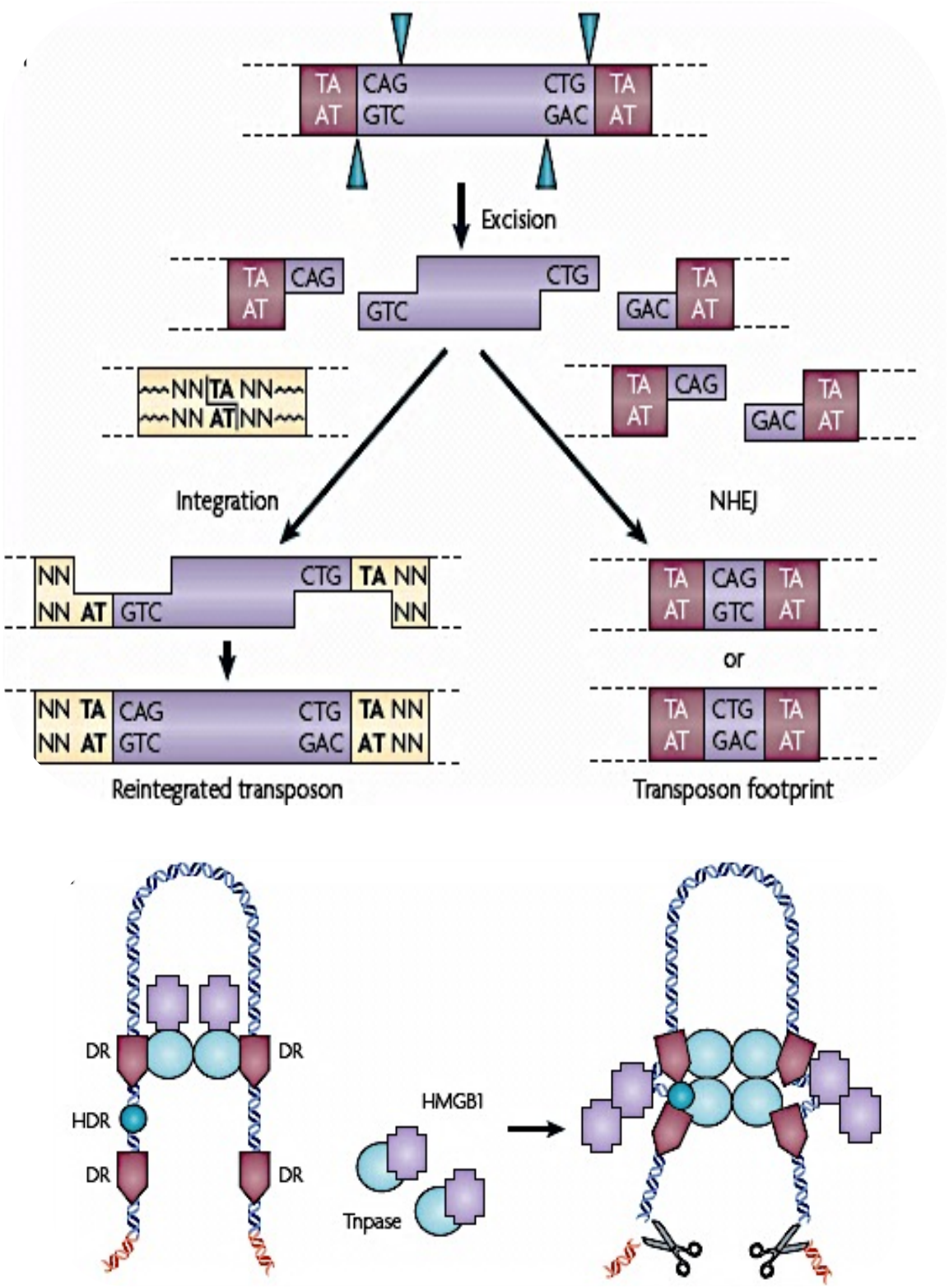

Figure 1-3 Cont.

advantage of the host's own receptor-mediated transport machinery. An AT-hook motif, which is made of two HTH motifs connected by a signature GRPR-like motif, has been found in *Sleeping Beauty* transposase as a functional domain responsible for DNA minor groove interaction. The catalytic domain of *Sleeping Beauty* transposase, which is responsible for DNA breakage and joining reactions, is characterized by conserved amino acid motif, the DDE motif (Figure 1-3). The DDE motif has been found in a large group of recombinases such as bacterial IS element transposases, retrotransposon/retroviral integrases, and RAG1 immunoglobin gene recombinase. It has also been suggested that the catalytic domain of DDE recombinases is involved in mediating interaction with the target DNA (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

The transposition process can be divided into four major phases: 1) transposase binding to the specific binding sites in the transposon DRs; 2) synaptic complex formation in which the two ends of the transposable elements are paired and held together by the transposase subunits; 3) excision from the donor site; 4) reinsertion at a new target site (Figure 1-3). Importantly, multiple layers of the regulatory molecular "checkpoints" are installed in each phase of the transposition process for *Sleeping Beauty* (or *piggyBac*) to mobilize its own transposable elements (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

One checkpoint is specificity of DNA binding by the *Sleeping Beauty* transposase. The PAI subdomain of *Sleeping Beauty* transposase recognizes the binding site of 3'DR/IR in base-specific manner whereas the RED subdomain interacts with the specific sequence of 5'DR/IR. Additionally, the PAI subdomain interacts with the HDR motif in

the left IR of *Sleeping Beauty* transposable element, and facilitates protein-protein interaction between other transposase subunits. As a consequence, the PAI domain has three critical regulatory roles in the first phase of transposition, such as interactions with both the DRs and the HDR motif, and transposase oligomerization. The zebra fish Tdr1 element is highly related to *Sleeping Beauty* transposons; however, the *Sleeping Beauty* transposase cannot mobilize the Tdr1 transposable element. It has been demonstrated that major differences lies in the 5'half of the DRs between the Tdr1 and *Sleeping Beauty* transposons. As the mediator of 5'binding specificity, the RED subdomain recognizes the difference in the 5'DRs from those of other transposable elements and prevents mobilization of the closely related transposable elements (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

The other checkpoint is the orders of *Sleeping Beauty* synaptic complex formation that are orchestrated by the high mobility protein HMGB1. The synaptic complex assembly is the process where the two ends of the *Sleeping Beauty* transposable element are held together by the *Sleeping Beauty* transposases (Figure 1-3). HMGB1 is non-histone nuclear protein that is associated with the eukaryotic chromatin and is capable of bending DNA. It has been shown that HMGB1 is required for effective *Sleeping Beauty* transposition in mammalian cells. Transposition efficiency of *Sleeping Beauty* transposons showed significant reduction in the cells from HMGB1-mutant mice. Furthermore, overexpression of HMGB1 promotes significant transposition efficiency in the normal mouse cells, which suggests that HMGB1 is a limiting factor of transposition. Additionally, HMGB1 was found to bind *Sleeping Beauty* transposase in vivo and assemble a ternary complex with the transposase and transposon DNA, which indicated

that the transposase might actively recruit HMGB1 to transposon DNA through protein-protein interaction. It has been reported that *Sleeping Beauty* transposase preferentially interacts with the inner DRs within the transposon inverted repeats. HMGB1 was shown to facilitate transposase binding more toward the inner DRs than the outer DRs. This evidence points out that HMGB1 ensures the orderly assembly of the catalytically ready synaptic complex during the early phase step of transposition. In fact, *Sleeping Beauty* transposase cannot achieve the transposition when the inner binding sites are swapped with the outer binding sites. Combined together, the ordered synaptic formation appears to regulate the cleavage of outer sites of DNA transposons, which occurs only when all the prerequisites have been met (Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

During the excision phase, the cuts at the ends of the *Sleeping Beauty* transposon are staggered inward by three nucleotides, which generate 3' overhangs and leave the nucleotides of the terminal inverted repeats (IRs) at the donor site. This creates the 3-bp transposon footprints upon the repair of double-strand DNA breaks of the broken DNA ends. In contrast, *piggyBac* transposase leaves no footprint during the excision of the *piggyBac* transposon from the donor site. *Sleeping Beauty* transposons integrate relatively randomly; however, *Sleeping Beauty* transposons do not utilize all available sites in the genome with equal frequencies (Handler 2002; Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010). Vigdal and colleagues (2002) performed in vitro transposition assay in human Hela cells to analyze insertion profile of *Sleeping Beauty* transposons. They showed that no clear preference was detected for any chromosome or for specific subchromosomal regions although some chromosomes were

hit more often than others. The results indicated that most chromosomes could be good targets for transposition. Vigdal and colleagues (2002) also found 35% of transposition events took place in transcribed regions. It has been suggested that approximately one third of the human genome is transcribed; therefore, this frequency of gene transcription indicates no preference for or against insertion into genes (International Human Genome Sequencing Consortium 2001). Furthermore, *Sleeping Beauty* transposon preferentially inserts at TA dinucleotides. Further analysis revealed that *Slepping Beauty* preferred a short palindromic AT-repeat: ATA**TAT**AT, where the central underlined TA is the insertion site. In contrast, *piggyBac* transposons preferentially integreate at tetranucleotide TTAA sites (Handler 2002; Mates et al., 2007; Izsvak & Ivics 2003; Carlson & Largaespada, 2005; Copeland & Jenkins 2010).

Ding and colleagues (2005) first demonstrated that the *piggyBac* transposon was able to effectively perform insertional mutagenesis in the mammalian cells and in mice (Ding et al., 2005). Following the success of Ding and colleagues, several laboratories have made further improvement of the *piggyBac* transposon system as the effective mutagenesis tool (Wu et al., 2007, Cadinanos and Bradley 2007; Roland et al., 2010; Kaji et al., 2010; Woltjen et al., 2010; Yusa et al., 2009). For instance, Bradley and colleagues have identified the coding sequence of *piggyBac* transposase, which optimize the transposition efficiency. Bradley and colleagues engineered inducible *piggyBac* transposase by fusing the optimized *piggyBac* transposase with the modified estrogen ligand-binding domain (ERT2), and showed the efficiency of this inducible system in mice (Cadinanos and Bradley 2007). Because of the high integration efficiency of *piggyBac* transposons, the *piggyBac* transposon system has been used as the effective

transgene deliver system of iPSC formation (Kaji et al., 2010; Woltjen et al., 2010). Since Bradley and collegaues have demonstrated that the integrated *piggyBac* transposons could be excised cleanly out of the host genome (Yusa et al., 2009), the *piggyBac* transposon system holds great promise to be further improved as a safer and effective method to generate patient-specific iPSCs. Collectively, DNA transposon systems such as *piggybac* transposons are a powerful genetic tool to study gene functions both in vitro and vivo.

References

Abeliovich A and Doege, C. (2009). Reprogramming therapeutics: iPS cell prospects for neurodegenerative disease. Neuron 61, 337-339.

Anokye-Danso, F., Trivedi, C. M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P. J., Epstein, J. A. and Morrisey, E. E. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell, 8, 376-388.

Araki, K., Araki, M. and Yamamura, K. (2006). Negative selection with the Diphtheria toxin A fragment gene improves frequency of Cre-mediated cassette exchange in ES cells. J Biochem., 140, 793-798.

Banito, A., Rashid, S. T., Acosta, J. C., Li, S., Pereira, C. F., Geti, I., Pinho, S., Silva, J. C., Azuara, V., Walsh, M., Vallier, L. and Gil, J. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev., 23, 2134-2139.

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H. and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell, 2, 151-159.

Brocard, J., Feil, R., Chambon, P. and Metzger, D. (1998). A chimeric Cre recombinase inducible by synthetic , but not natural ligands of the glucocorticoid receptor. Nucleic Acidas Res., 26, 4086-4090.

Brocard, J., Warot, X., Wendling, O., Messaddeo, N., Vonesch, J-L., Chambon, P. and Metzger, D. (1997). Spatio-temporally controlled site-specific somatic mutagenesis in the mouse. Proc Natl Acad Sci U.S.A., 94, 14559-14563.

Capecchi, M. R. (2001). Generating mice with targeted mutations. Nature Med., 7, 1086-1090.

Capechhi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet., 6, 507-512.

Cadinanos, J. and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. Nucleic Acids Res., 35, e87.

Carlson, C. M. and Largaespada, D. A. (2005). Insertional mutagenesis in mice: new perspective and tools. Nat Rev Genet., 6, 568-580.

Chai, Y., Jiang, X., Ito, Y., Bringas, P., Han, J., Rowitch, D. H., Soriano, P, McMahon, A. P. and Sucov, H. M. (2000). Fate of mammalian cranial neural crest during tooth and mandibular morphogenesis. Development, 127, 1671-1679.

Chin, M. H., Mason, M. J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., Khvorostov, I., Ott, V., Grunstein, M., Lavon, N., Benvenisty, N., Croce, CM., Clark, A. T., Baxter, T., Pyle, A. D., Teitell, M. A., Pelegrini, M., Plath, K. and Lowry, W. E. (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell, 5, 111-123.

Cho, HJ., Lee, C. S., Kwon, Y. W., Peak, J. S., Lee, S. H., Hur, J., Lee, E. J., Roth, T. Y., Chu, I. S., Leem, S. H., Kang, H. J., Park, Y. B. and Kim, H. S. (2010). Induction of pluripotent stem cells from adult somatic cells by protein-based reprogramming without genetic manipulation. Blood, 116, 386-395.

Cooley, L., Kelley, R. and Spradling, A. (1988). Insertional mutagenesis of the Drosophila genome with single P element. Science, 239, 1121-1128.

Copeland N. G. and Jenkins, N. A. (2010). Harnessing transposons for cancer gene discovery. Nat Rev Cancer, 10, 696-706.

Dimos, J. T., Rodolfa, K. T., Niakan, K. K., Weisenthal, L. M., Mitsumoto, H., Chung, W., Croft, G. F., Saphier, G., Leibel, R., Goland, R., Wichterle., H, Henderson, C. E. and Eggan, K. (2008). Induced pluripotent stem cells generated from patient with ALS can be differentiated into motor neurons. Science, 321, 1218-1221.

Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., Miller, J., Schlaeger, T., Daley, G. Q. and Feinberg, A. P. (2009). Differential methylation of tissue-and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet., 41, 1350-1355.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y. and Xu, T. (2005). Efficent transposition of the *piggyBac* (PB) transposon in mammalian cells and mice. Cell, 122, 473-483.

Ebert, A. D., Yu, J., Rose, F. F., Mattis, V. B., Lorson C. L., Thomson, J. A. and Svendsen, C. N. (2009). Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature, 457, 277-280.

Eminli, S., Utikal, J., Arnold, K., Jaenisch, R. and Hochedlinger, K. (2008). Reprogramming of neural progenitors cells into induced pluripotent stem cells in the absence of exogenous *Sox2* expression. Stem Cells, 26, 2467-2474.

Eminli, S., Foudi, A., Stadtfeld, M., Maherali, N., Ahfeldt, T., Mostoslavsky, G., Hock, H. and Hochedlinger, K. (2009). Differentiation stage determines potential of hemotopoietic cells for reprogramming into induced pluripotent stem cells. Nat Genet., 41, 968-976.

Epstein, J. A., Li, J., Lang, D., Chen, F., Brown, C. B., Jin, F., Lu, M. M., Thomas, M., Liu, E., Wessels, A. and Lo, C.W. (2000). Migration of cardiac neural crest cells in Splotch embryos. Development, 127, 1869-1878.

Esteban, M. A., Xu, J., Yang, J., Peng, M., Qin, D., Li, W., Jiang, Z., Chen, J., Deng, K., Zhong, M., Cai, J., Lai, L. and Pei, D. (2009). Generation of induced pluripotent stem cell lines from Tibetan miniature pig. J Biol Chem., 284, 17634-17640.

Feil, R. (2007). Conditional somatic mutagenesis in the mouse using site-specific recombinases. H.E.P., 178, 3-28.

Feil, R., Brocard, J., Mascrez, B., LeMur, M., Metzger, D. and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proc Natl Acad Sci U.S.A., 93, 10887-10890.

Feil, R., Wanger, J., Metzger, D. and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. Biochem Biophys Res Commun., 237, 191-197.

Feng, B., Ng, J. H., Heng, J. C. and Ng, H. H. (2010). Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. Cell Stem Cell, 4, 301-312.

Folger, K. R., Wong, E. A., Wahl, G. and Capecchi, M. R. (1982). Patterns of integrated DNA microinjected into cultured cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol Cell Biol., 2, 1372-1387.

Fusaki, N., Ban, N., Nishiyama, A., Saeki, K. and Hasegawa, M. (2009). Efficienct induction of transgene-free human pluripotent stem cells using a vector based on Sendai

virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad, 85, 348-362.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H. and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cell using cell type-specific gene targeting. Science, 265, 103-106.

Gunaseeli, I., Doss, M. X., Antzelevitch, C., Hescheler, J. and Sachinidis, A. (2010). Induced pluripotent stem cells as a model for accelerated patient-disease-specific drug discovery. Curr Med Chem., 17, 759-766.

Koch, P., Kokaia, Z., Lindvall, O. and Brüstle, O. (2009). Emerging concepts in neural stem cell research: autologous repair and cell-based disease modeling. Lancet Neurol., 8, 819-829.

Handler A. M. (2002). Use of the piggyBac transposon for germ-line transformation of insect. Insect Bioche Mol Biol., 36, 1211-1220.

Hanna, J., Wernig, M., Markoulaki, S., Sun, C W., Meissner, A., Cassady, J. P., Beard, C., Brambrink, T., Wu, L. C., Townes, T. M. and Jaenisch, R. (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science, 318, 1920-1923.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B. W., Beard, C., Wernig, M., Creyghton, M. P., Steine, E. J., Cassady, J. P., Foreman, R., Lengner, C. J., Dausman, J. A. Jaenisch, R. (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell, 133,250-264.

Hanna, J., Saha, K. and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell, 143, 508-524.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C. J., Creyghton, M. P., van Oudenaarden, A. and Jaenisch, R. (2009a). Direct cell reprogramming is stochastic process amenable to acceleration. Nature, 462, 595-601.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A. W., Cassady, J. P., Staerk, J., Carey, B. W., Lengner, C. J., Foreman, R., Love, J., Gao, Q., Kim, J. and Jaenisch, R. (2009b). Metastable pluripotent states in NOD-mouse-derived ESCs. Cell Stem Cell, 4, 513-524.

Hartl, D. L., Lozovskaya, E. R., Nurminsky, D. I. and Lohe, A. R. (1997). What restricts the activity of mariner-like transposable elements? Trends Genet., 13, 197-201.

Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L. E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J. A., Ecker, J. R.

and Ren, B. (2010). Distinct epigenetic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell, 6, 479-491.

Herrera, P. L. (2000). Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development, 127, 2317-2322.

Hoess, R. H. and Amremski, K. (1990). The Cre-lox recombination system. In Eckstein, F. and Lilley, D. M. J. Nucleic acids and molecular biology, vol 4. Springer-Verlag, NY, 99-109.

Hochedlinger, K and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. Development, 136, 509-523.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K. and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p-53-p21 pathway. Nature, 460, 1132-1135.

Ichida, J. K., Blanchard, J., Lam, K., Son, E. Y., Chung, J. E., Egli, D., Loh, K. M., Carter, A. C., Di Giorgio, F. P., Koszka, K., Huangfu, D., Akutsu, H., Liu, D. R., Rubin, L. L. and Eggan, K. (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell Stem Cell, 5, 491-503.

International Human Genome Sequencing Consortium. (2001). Initial sequencing analysis of the human genome. Nature, 409, 860-921.

Ivics, Z., Kaufman, C. D., Zayed, H., Miskey, C., Walisko, O. and Izsvák, Z. (2004). The Sleeping Beauty transposable element: Evolution, regulation and genetic application. Curr Issues Mol Biol., 6, 43-56.

Izsvak, Z. and Ivics, Z. (2004). Sleeping Beauty transposition: biology and application for molecular therapy. Mol Ther., 9, 147-156.

Jacob, J. and Baltimore, D. (1999). Modeling T-cell memory by genetic marking of memory T cells in vivo. Nature, 399, 593-597.

Jaenisch, R. and Young R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell, 132, 567-582.

Jia, F., Wilson, K. D., Sun, N., Gupta, D. M., Huang, M., Li, Z., Panetta, N. J., Chen, Z. Y., Robbins, R. C., Kay, M. A., Longaker, M. T. and Wu, J. C. (2010). A nonviral minicircle vector for deriving human iPS cells. Nat Methods, 7, 197-199.

Jiang, X., Rowitch, D. H., Soriano, P., McMahon, A. P. and Sucov, H. M. (2000). Fate of mammalian cardiac neural crest. Development, 127, 1607-1617.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P. and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature, 458, 771-775.

Kellendonk, C., Tronche, F., Managha, A-P., Angrand, P-O., Stewart, F. and Schutz, G. (1996). Regulation of Cre recombinase activity by the synthetic steroid RU 486. Nucleic Acids Res, 24, 1404-1411.

Kellondonk, C., Troche, F., Casanova, E., Analag, K., Opherk, C. and Schutz, G. (1999). Inducible site-specific recombination in the brain. J Mol Biol, 285, 175-182.

Koller, B. H. and Smithies, O. (1992). Altering genes in animals by gene targeting. Annu Rev Immunol, 10, 705-730.

Kuhn, R., Schwenk, F., Aguet, M. and Rajewsky, K. (1995). Inducible gene targeting in mice. Science, 269, 1427-1429.

Kwan, K-M. (2002). Conditional alleles in mice: practical considerations for tissue-specific knockouts. Genesis, 32, 49-62.

Lasko, M., Sauer, B., Mosinger, B., Lee, E. J., Manning, R. W., Yu, S. H., Mulder, K. L. and Weatphal, H. (1992). Targeted oncogene activation by site-specific recombination in transgenic mice. Proc Natl Acad Sci U.S.A., 89, 6232-6236.

Lewandoski, M. (2002). Conditional control of gene expression in the mouse. Nat Rev Genet., 2,743-755.

Lewin, B. (2001). Genes VIII. Pearson Prentice Hall, NJ.

Li, H., Collado, M., Villasante, A., Strati, K., Ortega, S., Canamero, M., Blasco, M. A. Serrano, M. (2009). The Ink/Arf locus is a barrier for cell reprogramming. Nature, 460, 1136-1139.

Li, W. and Ding, S. (2010). Small molecules that modulate embryonic stem cell fate and somatic cell reprogramming. Trends Pharmacol Sci., 31, 36-45.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., Qin, B., Xu, J., Li, W., Yang, J., Gan, Y., Qin, D., Feng, S., Song, H., Yang, D., Zhang, B., Zeng, L., Lai, L., Esteban, M. A. and Pei, D. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell, 7, 51-63.

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S. Y., Qin, J., Wong, J., Cooney, A. J., Liu, D. and Songyang, Z. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol., 10, 731-739.

Liao, J., Cui, C., Chen, S., Ren, J., Chen, J., Gao, Y., Li, H., Jia, N., Chend, L. and Xiao, H. (2009). Generation of induced pluripotent stem cell lines from adult rat cells. Cell Stem Cell, 4, 11-15.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, A., Ren, B. and Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature, 462, 315-322.

Liu, H., Zhu, F., Yong, J., Zhang, P., Hou, P., Li, H., Jiang, W., Cai, J., Liu, M., Cui, K., Qu, X., Xiang, T., Lu, D., Chi, X., Gao, G., Ji, W., Ding, M. and Deng H. (2008). Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. Cell Stem Cell, 3, 587-590.

Maherali, N., Sridharan R., Xia, W, Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M, Yachechoko, R., Tchieu, J. and Jaenisch, R, Plath, K. anf Hochedlinger, K. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell, 1, 55-70.

Maherali, N., Ahfeldt, T., Rigamanti, A., Utikal, J., Cowan, C. and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell, 26, 340-345.

Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988). Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature, 336, 348-352.

Marión, R. M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M. and Blasco, M. A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature, 460, 1149-1153.

Mátés, L., Izsvák, Z. and Ivics, Z. (2007). Technology transfer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives. Genome Biol., 8, S1.

Metzger, D., Clifford, J., Chiba, H. and Chambon, P. (1995). Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. Proc Natl Acad Sci U.S.A., 92, 6991-6995.

Mikkelsen, T. S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B. E., Jaenisch, R., Lander, E. S. and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature, 454, 49-55.

Okita, K., Ichisaka T. and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature, 448, 313-317.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. and Yamanaka, S. (2008). Generation of mouse pluripotent stem cells without viral vectors. Nature, 448, 313-317.

Polo, J. M., Liu, S., Figueroa, M. E., Kulalert, W., Eminli, S., Tan, K. Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., Natesan, S., Wagers, A. J., Melnick, A., Evans, T. and Hochedlinger, K. (2010). Cell types of orgin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol., 28, 845-855.

Plasterk, R. H. (1996). The Tc1/mariner transposon family. Curr Top Microbiol Immunol., 204, 125-143.

Ryding, A. D. S., Shrp, M. G. F. and Mullins, J. J. (2001). Conditional transgenic technologies. J Endocrinol., 171, 1-14.

Saha K. and Jaenisch, R. (2009). Teachnical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell, 5, 584-595.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H. K., Beyer, T. A., Datti, A., Woltjen, K., Nagy, A. and Wrana, J. L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell, 7, 64-77.

Schwenk, F., Kuhn, R., Angran, P-O., Rajewsky, K. and Stewart, F. (1998). Temporally and spatially regulated somatic mutagenesis in mice. Nucleic Acids Res., 26, 1527-1432.

Silva, J., Chambers, I., Pollard, S. and Smith, A. (2006). Nanog promotes transfer of pluriopotency after cell fusion. Nature, 441, 997-1001.

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell, 138, 723-737.

Sommer, C. A., Sommer, A. G., Longmire, T. A., Christodoulou, C., Thomas, D. D., Gostissa, M., Alt, F. W., Murphy, G. J., Kotton, D. N. and Mostoslavsky, G. (2010). Excision of reprogramming transgene improves the differentiation potential of iPS cells generated with a single excisable vector. Stem Cells, 28, 64-74.

Sommer, C. A., Stadtfeld, M., Murphy, G. J., Hochedlinger, K., Kotton, D. N. and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a lentiviral stem cell cassette. Stem Cells, 27, 543-549.

Sommer, C. A., Sommer, A. G., Longmire, T. A., Christodoulou, C., Thomas, D. D., Gostissa, M., Alt, F. W., Murphy, G. J., Kotton, D. N. and Mostoslavsky, G. (2010).

Excision of reprogramming transgene improves the differentiation potential of iPS cells generated with a single excisable vector. Stem Cells, 28, 64-74.

Sridharan, R., Tchieu, J., Mason, M. J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q. and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell, 136, 364-377.

Stadtfeld, M and Hochedlinger, K (2010). Induced pluripotency: history, mechanisms, applications. Genes Dev., 24, 2239-2263.

Stadtfeld, M., Brennand, K. and Hochedlinger, K. (2008a). Reprogramming of pancreatic β cells into induced pluripotent stem cells. Curr Biol., 18, 890-894.

Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. and Hochedlinger, K. (2008b). Induced pluripotent stem cells generated without viral integration. Science, 322, 945-949.

Stadtfeld, M., Maherali, N., Breault, D. T. and Hochedlinger, K. (2008c). Defining molecular cornerstones during fibroblasts to iPS cell reprogramming in mouse. Cell Stem Cell, 2, 230-240.

Szybaiski, W. (1992). Use of the HPRT gene and the HAT selection technique in DNA-mediated transformation of mammalian cells: first steps toward developing hybridoma techniques and gene therapy. Bioessays, 14, 495-500.

Takahashi, K and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblasts cultures by defined factors. Cell, 126, 663-676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 131, 861-872.

Thomas, K. R. and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell, 51, 503-512.

Tsubooka, N., Ichisaka, T., Okita, K., Takahashi, K., Nakagawa, M. and Yamanaka, S. (2009). Roles of Sal4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. Genes Cells., 14, 683-694.

Utikal, J., Polo, J.M., Stastfeld M., Maherali, N., Kulalert, W., Qalsh, R.M., Khalil, A., Rheinwald, J.G. and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature, 460, 1145-1148.

Vigdal, T. J., Kaufman, C. D., Izsvak, Z., Voytas, D. F. and Ivics, Z. (2002). Common physical properties of DNA affecting site selection of sleeping beauty and other Tc1/Mariner transposable elements. J Mol Biol., 323, 441-452.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W. and Orkin, S. H. (2006). A protein interaction nerwork for pluripotency of embryonic stem cells. Nature, 444, 364-368.

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y. H., Li, H., Lau, F., Ebina, W., Mandal, P. K., Smith, Z. D., Meissner, A., Daley, G. Q., Brack, A. S., Collins, J. J., Cowan, C., Schlaeger, T.M. and Rossi, D. J. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell, 7, 618-630.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E. and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature, 448, 318-324.

Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H. K. and Nagy, A. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature, 458, 766-770.

Wu, S., Ying, G., Wu, Q. and Capecchi, M. R. (2007). Toward simpler and faster genome-wide mutagenesis in mice. Nat Genet., 39, 922-930.

Xu, D., Alipio, Z., Fink, L. M., Adcock, D. M., Yang, J., Ward, D. C. and Ma, Y. (2009). Phenotypic correction of murine hemophilia A using an iPS cell-based therapy. Proc Natl Acad Sci U.S.A., 106, 808-813.

Yang, J., Chai, L., Fowles, T. C., Alipio, Z., Xu, D., Fink, L. M., Ward, D. C. and Ma, Y. (2008). Genome-wide analysis revealed Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. Proc Natl Acad Sci U.S.A., 105, 19756-19761.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I. and Thomson, J. A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science, 324, 797-801.

Yusa, K., Rad, R., Takeda, J. and Bradley, A. (2009). Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. Nat Methods, 6, 363-369.

Zhou, H. and Freed, C. R. (2009). Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells. Stem Cells, 27, 2667-2674.

Zhou, H., Wu, S., Joo, J. Y., Zhu, S., Han, D. W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., Siuzdak, G., Scholer, H. R., Duan, L. and Ding, S. (2009). Generation of induced pluripotent stem cells using recombinant proteins. Cell Stem Cell, 4, 381-384.

Zinyk, D. L., Mercer, E. H., Harris, E., Anderson, D. J. and Joyner, A. L. (1998). Fate mapping of the mouse midbrain-hindbrain constriction using a site-specific recombination system. Curr Biol., 8, 665-668.

CHAPTER 2

CHARACTERIZATION OF CANINE QUANTITATIVE TRAIT LOCUS (QTL) IN MOUSE BY CANINE iPS CELLS AND CHROMOSOME TRANSLOCATION TECHNOLOGY

Introduction

The dog (Canis familiaris) has been not only the best friend of the humans but also the most valuable work force for long time since its first domestication, which is estimated to be between 15,000 and 150,000 years ago (Ostrander et al, 2005; Wayne & Ostrander, 2005: Ostrander & Wayne; 2005; Sutter & Ostrander, 2004). Over centuries the dog has served us shepherd, guide, hunter, protector and a model organism for medical research. In order to meet needs and demands in our human lives, we desire to create dogs with particular physical and behavioral traits or special skills (Ostrander et al, 2005; Wayne & Ostrander, 2005; Ostrander & Wayne, 2005; Sutter & Ostrander, 2004). For centuries, we have been practicing selective breeding to create a perfect breed meeting a certain phenotypic standard (e.g., temperament and sharpness) (Ostrander et al., 2005; Wayne & Ostrander, 2005; Ostrander & Wayne, 2005; Sutter & Ostrander, 2004). The chief purpose of the selective breeding is to create dogs that are physically "fit" to specific tasks. For instance, the long-bodied dachshunds are made for hunting badgers underground. The heavy-coated komondors are developed to guard large herds

of animals on the open plains. The web-footed newfoundlands are made to be natural and powerful swimmers (Ostrander et al, 2005; Wayne & Ostrander, 2005;Ostrander & Wayne; 2005; Sutter & Ostrander, 2004). As a result, the selective breeding practice of dogs has created a broad variety of phenotypic traits with different heights, weights, bone structures and hair coats (Ostrander et al, 2005; Wayne & Ostrander, 2005; Ostrander & Wayne, 2005; Sutter & Ostrander, 2004). In point of fact, the domestic dogs are known to be the most morphologically and diverse mammalian species on earth, and almost all combinations of physical traits have been generated and fixed within specific breeds, which makes domestic dogs an exceptional model for studying morphological variations in humans (Ostrander et al., 2005; Lark et al., 2006).

Lark and his colleagues take advantage of the morphological diversity of the domestic dogs and examine the genetic basis for skeletal variation in Portuguese water dogs (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). Among the pure breeds of domestic dogs, the Portuguese water dog is a breed that is allowed large variation in skeletal size by the American Kennel Club, which makes this breed an attractive breed to study skeletal variation. Furthermore, this breed consists of a reasonable number of individuals (~10000) that were generated by a small founders population (33 dogs)(Ostrander et al., 2005; Lark et al., 2006). Because of generous support from owners, phenotypic and genotypic materials, which are X-rays and blood samples respectively, are easily corrected. The accuracy of the pedigree is tremendous so that individual dogs can be traced back to the founders (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). Finally, their primary assessment, which is the measurements of 12 morphological parameters (e.g., height at shoulder, length of tail,

length of body, height at hock, distance between eyes) measured by 300 owners with tape measures, revealed that morphological phenotypes segregate. Simulation based on these raw date could prove the evidence of sufficient heritability of different morphological/quantitative traits, which would lead to identification of the loci of those quantitative traits (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005).

Quantitative traits are also called complex traits because the traits are affected by multiple genetic and environmental factors. Regions in the genome where it harbors multiple genes contributing to a complex trait are known as quantitative trait loci (QTLs) (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). Quantitative traits such as metric traits, which are measurable on a continuous scale (e.g., height or weight), are distinguished from traits that look discrete such as round versus wrinkle pea seeds (Chase et al., 1999). In studies of the Lark's group, metrics of the canine skeleton obtained by direct measurement or from radiographs are used as the quantitative traits of interest (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). Principle component (PC) analysis was used to classify phenotypic variation into independent systems of correlated traits (Jackson, 1991; Hartl, 2001). Individual dogs have a value for every PC. Thus, PCs are phenotypes that are subject to genetic analysis and QTLs can be identified that inform these phenotypes (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). Although this analysis uses numerous skeletal metrics, which increases statistical power, the analysis focuses on the pattern of multiple metrics and will not capture independent variation of the metrics

of individual bones. Therefore, it enables the researchers to detect QTLs with pleiotropic effects preferentially (Lark et al., 2006; Chase et al., 2002; Carrier et al., 2005).

PC analyses in skeletal variations in Portuguese water dogs revealed that PCs individually promote both correlations within structures (e.g., within the skull or among the limb bones) and inverse correlations between structures (e.g., skull vs. limb bones). This finding is described as a trade-off that forms compromise between high production of power and energy-efficient speed (Lark et al., 2006; Chase et al., 2002; Carrier et al., 2005). For instance, a smaller head and large postcranial body is linked to speed, which is observed in the greyhound, and a large head and smaller postcranial body is associated with power, such as a pit bull. This sharp variation is described by one principle component. Furthermore, there is another PC characterizing independent component of sharp variation that influences metrics of length and width (Lark et al., 2006; Chase et al., 2002; Carrier et al., 2005). For instance, a long and thin limb bone is connected to speed (e.g., greyhounds), whereas a shorter and thicker limb bone is associated with power (e.g., pit bulls). Using phenotypic data from the radiograph corrected and polymorphic molecular makers (Simple Sequence Repeats or SSR), the group has identified a QTL that regulates the principle component involving a trade-off between length and width of limb bones. This QTL is associated with the SSR marker FH3585 on the canine chromosome (CFA) 12 (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). At this locus, specific haplotypes control both length and width of limb bones. D haplotye and E haplotype regulates both length and width of radius. D haplotype gives rise to a wider but shorter radius whereas E haplotype generates a thinner and longer radius. The effects of these haplotypes appear to be those of regulatory

gene(s) or element(s), which inhibit growth in one aspect while stimulating growth in another. These haplotypes are the ones that we are interested in charactering (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). The SSR marker FH3585 is located approximately 39Mb from the centromere of CFA12 whose total size is 76Mbs. Within 10Mbs around the marker FH3583, more detailed analyses of this particular QTL have been performed with additional SSR markers. Two markers along FH3585 showed the strong association with this QTL at approximately 35Mbs and 47Mbs from the centromere of CFA12, respectively. Lark and colleagues suggest that this 12Mbs region between two markers may contain gene(s) or regulatory element(s) responsible for this trade-off phenotype in the limb bone (Lark et al., 2006; Chase et al., 1999; Chase et al., 2002; Carrier et al., 2005). However, because the downstream of CFA12 has not been fully characterized yet, it has not ruled out the possible makers that may be associated with this particular QTL at the downstream of CFA 12. In fact, when the downstream of CFA 12 (i.e., the region after the 47Mb SSR marker) is surveyed, there are several candidates, such as different collagen genes, Ephrin receptors (Stadler et al., 2001), and transcription factor FoxO3 Kaufmann & Knochel, 1996: Lehmann, 2003) for instance, that may play an important role in bone formation.

Longitudinal growth of bone depends mainly on the rate of production of hypertrophic chondrocytes from the proliferating chrondrocytes in a process of endochondral ossification (Karsenty, 2003; Kronenberg, 2003). Endochondral ossification starts with the condensation of mesenchymal cells, which are adhered through the expression of adhesion molecules. The aggregated mesenchymal cells differentiate into chondrocytes, the primary cells type of cartilage. Expressions of the

adhesion molecules, N-cadherin and N-CAM, are induced by bone morphogenetic proteins (BMPs). N-cadherin and N-CAM appear to be important for the initiation and maintenance of mesenchymal condensation (Adams et al., 2007; Kobayashi et al., 2007; Wu et al., 2007). BMP is also responsible for inducting the expression of the transcription factor *Sox9,* which activates other transcription factors as well as the type II collage gene and the proteoglycan aggrecan (Adams et al., 2007; Kobayashi et al., 2007; Wu et al., 2007). The chondrocytes go through rapid proliferation to form the cartilage model of the bone. During their rapid proliferation, the chondrocytes produce and secrete a cartilage-specific extracellular matrix. Then, the condrocytes in the center of condensation stop dividing, increase their volume radically, and become hypertrophic chondrocytes. As the transformation of chondrocytes occurs, those cells change the synthesis of the cartilage-specific extracellular matrix by adding the collagen X, which enables the matrix to be mineralized by calcium phosphate. The hypertrophic chondrocytes secrete the angiogenesis factor, *VEGF*, and other factors to promote the invasion of blood vessels into the cartilage model (Karsenty, 2003; Kronenberg, 2003; Adams et al., 2007; Kobayashi et al., 2007; Wu et al., 2007). Hypertrophic chondrocytes lead adjacent perichondrila cells to become osteoblasts, which secrete a specific matrix to generate a bone collar. Then hypertrophic chondrocytes go through the apoptotic cell death. Osteoblasts use the cartilage matrix that was left behind as a scaffold to invade the cartilage mould accompanied by blood vessels. Eventually, osteoblasts put down a true bone matrix within it, which is known as primary spongiosa. While those activities of hypertrophic chondrocytes occur, the proliferation of chondrocytes continues in the mould and the mould therefore enlarges (Karsenty, 2003; Kronenberg, 2003; Adams et

al., 2007; Kobayashi et al., 2007; Wu et al., 2007). A fraction of these proliferating chondrocytes changes the morphology to a flattened, discoid shape and stack themselves up with a clear orientation that promotes the lengthening of the bone. As the bone elongates further, secondary ossification centers are set up when chondrocytes repeat the aforementioned process. In the long bone of limbs, it is the region where the condrocytes continue to proliferate between primary and secondary ossification centres, which are defined as the growth plate. At the top of the growth plate, some chondrocytes become round and slow down their proliferation. These cells are known as reserve chondrocytes, which serve as precursors for the flat proliferating columnar chondrocytes (Karsenty, 2003; Kronenberg, 2003; Adams et al., 2007; Kobayashi et al., 2007; Wu et al., 2007).

Numerous signaling pathways and transcription factors including a BMP signaling pathway and *Runx2* orchestrate the multiple events in the endochondral ossification. BMPs play multiple roles during bone formation. For example, the type 1 receptors *BMPR1A* and *BMPR1B* are expressed in embryonic mesenchyme and cartilage condensation, respectively (Adams et al., 2007; Kobayashi et al., 2007; Wu et al., 2007). When the BMP antagonist, *Noggin,* was expressed in early chick limbs, the mesenchyme failed to form the condensation (Pizette & Niswander, 2000). Furthermore, the enlarged cartilage primordial was observed in *Noggin*-/-mice (Brunet et al., 1998). Those studies indicate important roles of BMP signaling in formation of mesenchymal condensation. *BMP2*, *3, 4, 5*, and *7* are expressed in perichondrium. Expressions of *BMP 2* and *6* are observed in hypertrophic chondrocytes. *BMP* 7 also displays its expression in proliferating chondrocytes (Adams et al., 2007; Kobayashi et al., 2007; Wu et al., 2007). When BMPs were added to bone explants, proliferation of chondrocyates increased

whereas it inhibited with Noggin (Minina et al 2001, 2002). Furthermore, terminal differentiation of hypertrophic chondrocytes was either delayed or hastened by *BMP2* and *Noggin* treatment, respectively (Minina et al 2001, 2002). Finally, studies of the mouse short ear (se) locus, which harbors numerous inactivating mutations of *BMP5*, displayed significant alterations in the size, shape, and number of many small bone and cartilage elements with other skeletal structures unaffected. The signature phenotype of the se mouse carried is reduced size of external ears, which is caused by the defect in formation of mesenchymaal condensation (Kingsley et al., 1992; King et al., 1994).

A transcription factor *Runx2*, a formerly known as *Cbfa1*, appears to be the master regulator in bone formation (Schroeder et al., 2005; Otto et al., 1997; Komori et al., 1997; Inada et al., 1999). The expression pattern of *Runx2* changes over time. *Runx2* expression starts in the last condensation phase of chondrogenesis. The expression significantly decreases in proliferating chondrocytes. Then, the expression substantially increases in prehypertrophic and hypertrophic chondrocytes. Also, high expression of *Runx2* is identified in perichondrial cells and osteoblasts (Takeda et al., 2001). *Runx2* mutant mice lack osteoblasts and also display defects in chondrocyte maturation. The bone phenotypes observed in the mutant are either loss or significant decrease of hypertrophic chondrocytes and failure of matrix mineralization by the existing hypertrophic chondrocyte (Schroeder et al., 2005; Otto et al., 1997; Komori et al., 1997; Inada et al., 1999). Moreover, the mutant mice showed an absence of the expressions of osteopontin and matrix metalloproteinase, which are normally expressed in late hypertrophic chondrocytes (Schroeder et al., 2005; Otto et al., 1997; Komori et al., 1997; Inada et al., 1999). The transgene expression of *Runx2* is capable of rescuing the

aforementioned phenotypes. Overexpression of *Runx2* seeded up the hypertrophy of normal chindrocytes and then stimulated hypertrophy and additional bone formation in cartilage that never usually experience hypertrophy, including tracheal rings. The results mentioned above suggest that *Runx2* regulates differentiation of hypertrophic choondrocytes and osteoblasts (Schroeder et al., 2005; Otto et al., 1997; Komori et al., 1997; Inada et al., 1999).

Significant roles of *Runx2* and BMP signaling pathway in bone formation were described above. Interestingly, both *BMP5* and *Runx2* are found at distant upstream (9Mb and 19Mb, respectively) of the QTL of interest in canine chromosome12. It has been reported that developmental genes and transcription factors including *BMP5* utilize distant regulatory elements to achieve its tissue-specific expressions (DiLeone et al, 1998, 2000).

In complex genomes such as those of human and mouse, many genes require a *cis*-acting element in genome to ensure temporally and spatially correct activation/ repression of gene expression. Those *cis*-regulatory elements are defined as enhancer or repressor when the element activates or silences gene expression in a specific place (i.e., cell/tissue types) and certain time. Enhancers and repressors can be found afar upstream or downstream of their core transcription units. Those regulatory elements are also embedded within intron of genes. Sometimes, these are located within an intron of a gene that is functionally unrelated to the regulated gene. To date, the distance between the *cis*-regulatory elements and its core transcription unit can be extended as much as 1Mb (Dekker, 2008; Li et al., 2006; Kadauke & Blobel, 2008; West & Fraser, 2005; Bondarenk et al., 2003).

How a remote enhancer/repressor regulates its target gene remains unknown. However, it has been hypothesized that widely spaced regulatory elements (i.e., enhancers and repressors) in complex genomes can communicate with the target genes by conformation changes in chromosome that bring the remote regulatory elements and their target genes in proximity (Dekker, 2008; Li et al., 2006; Kadauke & Blobel, 2008; West & Fraser, 2005; Bondarenk et al., 2003). It has been suggested that histone acetylation and other histone modifications modulate conformation changes in chromosomes, which makes chromatin form a loop. Histone acetylation and other modifications seem to play key roles in determining the probability of interactions between distant regulatory elements and their target genes. Several hypotheses of long-range communication between remote regulatory elements and their cognate genes have been put forward (Dekker, 2008; Li et al., 2006; Kadauke & Blobel, 2008; West & Fraser, 2005; Bondarenk et al., 2003; Miele & Dekker, 2008). First of all, the simplest and the most attractive model is chromatin looping between the enhancer/repressor and promoters. As described earlier, histone modifications lead to exposure of the preferential looping sites in chromatin, which determines the chance of interaction between remote enhancers/repressors and its target genes. Second of all, according to a tracking or scanning model, once an enhancer recruits the transcription-activating complex, the complex linearly moves along chromatin until the complex meets its cognate promoter. This model does not involve the conformation changes in chromatin such as the formation of loop in order for the enhancers to interact its distant target promoters. A third model is the facilitated tracking model, which combines aspects of chromatin looping and tracking model together. According to the facilitated tracking model, an

enhancer and its transcription-activating complex travel along the chromatin until they encounter the target promoter. While the transcription-activating complex with the enhancer migrates toward its target promoter, the chromatin fiber that the complex travels along is pushed out through the complex with the enhancer and creates a loop. As the complex is skating along the chromatin, the loop gets enlarged. Lastly, the linking model suggests that facilitator proteins that bind to a chromatin fiber between an enhancer and its target gene make a chain of facilitator protein, which may define the activated domain and mediate the communication between the enhancer and its cognate promoter (Dekker, 2008; Li et al., 2006; Kadauke & Blobel, 2008; West & Fraser, 2005; Bondarenk et al., 2003; Miele & Dekker, 2008).

Although it remains unknown which model of communication remote regulatory elements utilize to regulate their target genes, many distant regulators have been identified over the years. Examples are the transcription factors, the winged helix/forkhead (*FOX*) genes and the bone morphogenetic proteins (BMPs). It has been suggested that a long-range control of the winged helix/forkhead (*FOX*) genes plays a critical role in ocular diseases among humans (Kleinjan & van Heyningen, 2005). The *FOX* genes are a large family of transcription factors characterized by 100-amino-acid winged helix DNA-binding domain. This family of transcription factors is involved in a wide range of developmental pathways including eye development. *FOXC1*, previously known as *FKHL7*, is a part in a cluster of forkhead genes on chromosome 6p25 (Kleinjan & van Heyningen, 2005; Lehmann, 2003; Kaufmann & Knochel, 1996). Mutations cause ocular malformations with iris or extra-ocular abnormalities and glaucoma (iridogoniodysgenesis type 1, IRID1). In addition to ocular malformations with iris and

extra-ocular abnormality, the cardiac, renal, meningeal, skeletal and somite developmental defects in *Foxc1* knockout mice and the spontaneous *Foxc1* mutant, congenital hydrocephalus. Those mutants demonstrate the diverse involvements of *FOX* genes in multiple organ developments. Whether a number of copies of *FOXC1* increases or decreases due to 6p25 segmental duplication and deletion, the phenotypes are very similar, which suggests that the precise gene dosage is crucial for normal eye development (Kleinjan & van Heyningen, 2005; Lehmann, 2003; Kaufmann & Knochel, 1996). Besides the intragenic point mutations and dosage effects for the *FOXC1* gene in patients with IRID1, it has been reported that a patient suffering from primary congenital glaucoma carries a balanced translocation mapped 25 kb from the gene. Also, it has been discovered that a patient with glaucoma and autosomal dominant iris anomaly was identified to harbor an interstitial deletion of 6p24-p25. This deletion was approximately located at distance of 1200kb from the *FOXC1* locus. Those cases indicate that distant enhancers may regulate the tissue-specific expression of *FOXC1* (Kleinjan & van Heyningen, 2005; Dekker, 2008; Li et al., 2006; Kadauke & Blobel, 2008; West & Fraser, 2005; Bondarenk et al., 2003; Miele & Dekker, 2008).

FOXC2, another member of *FOXC* family, may possibly be regulated by the distant regulatory elements. In development, *Foxc2* plays a role in early mesoderm patterning. *Foxc2* knockout mice die of the skeletal genitourinary and cardiovascular defects, which is similar to *Foxc1* homozygous knockout mice (Kleinjan & van Heyningen, 2005; Lehmann, 2003; Kaufmann & Knochel, 1996). In humans, *FOXC2* is located at chromosome 16q24. *FOXC2* mutations have been associated with lymphedema distichiasis (LD), which is an autosomal dominant disorder described by

lymphedema affecting the limbs and double rows of eyelashes (distichiasis). It has been reported that a patient suffering from neonatal lymphedema carries a translocation t(Y;16)(q12;q24.3) whose breakpoint is approximately at 120kb away from 3'end of *FOXC2* gene. The translocation apparently does not disrupt either genes on chromosome 16 and any candidate gene on the Y chromosome (Kleinjan & van Heyningen, 2005; Dekker, 2008; Li et al., 2006; Kadauke & Blobel, 2008; West & Fraser, 2005; Bondarenk et al., 2003). This strongly suggests a possible presence of tissue-specific regulatory element of *FOXC2* at this particular breakpoint.

The bone morphogenetic proteins (BMPs) are a large subclass of the transforming growth factor-β (TGF-β) family of secreted signaling molecules (Kingsley, 1994). It has been shown that BMPs play a critical role in embryonic development such as dorsal-ventral and left-right axis formation, mesenchymal-epithelial interaction and differentiation of many tissues including lung, gut, kidney, hair, teeth, cartilage and bone (Kingsley, 1994). Lines of evidences indicate that some of BMP family members utilize distant *cis*-regulatory elements to accomplish precise spatiotemporal transcription (DiLeone, et al., 1998, 2000; Chandler et al., 2007). Studies of the Drosophila decapentaplegic (dpp) have illustrated that several distinct and sometimes distant sequences specifically control dpp gene expression (Manak et l., 1994; Huang et al., 1993). Expression of the gene in the midgut is regulated by sequences positioned 5'of the gene. Early embryonic expression is controlled by sequences in the intron. Some of the dpp enhancers are placed more that 30kb away from the promoter in the considerably condensed fly genome (Manak et al., 1994).

In vertebrate, a spatial and temporal transcription regulation of BMPs has been extensively examined. Among those, distant control elements of *Bmp2* and *Bmp5* have been identified (DiLeone, et al., 1998, 2000; Chandler et al., 2007). For studies of both *Bmp2* and Bmp5, LacZ reporter transgene approach was used to survey large regions of Bmp genes (106kb of 3'genomic DNA for the mouse Bmp5 and approximately 400kb surrounding the mouse *Bmp2*) and identify DNA sequences that can drive the expression of a reporter gene in *Bmp* expression domains. Regulatory mutations and transgenic report studies for both *Bmp2* and *Bmp5* strongly indicated that a number of separated DNA sequences were positioned at separate genomic regions flanking either side of those genes (both 5' and 3' genomic regions for *Bpm2* and 3' genomic region for *Bmp5*) and those sequences are capable of regulating expressions of Bmp genes at specific anatomical locations in their designated tissues (DiLeone, et al., 1998, 2000; Chandler et al., 2007). *cis*-regulatory sequences that regulate *Bmp2* expression in osteoblast progenitors were discovered in the distant 3'flanking region of *Bmp2*. *Bmp2* expression controlled by the potential osteoblast progenitor regulatory element closely follows expression of *RUNX2* gene, the vital osteoblast factor. This potential osteoblast progenitor regulator identified as well-conserved 656bp sequences that were located 156.3kb from the *Bmp2* promoter (Chandler et al., 2007). For regulation of *Bmp5* expression, control elements in a distant 3' region of *Bmp5* gene, including manubrium, genital tubercle, thyroid cartilage, and lung mesenchyme, showed enormous specificity in the regulation of *Bmp5* expression at specific anatomical locations in specific tissues (DiLeone, et al., 1998, 2000). For instance, in the skeleton, regulatory sequences identified in 3' genomic region of *Bmp5* could drive a restricted reporter expression in the

upper region of a single bone at the top of the sternum (rostral manubrium). Normal expressions of *Bmp5* were seen in all the skeletal structures (DiLeone, et al., 1998, 2000). However, when a series of mutations were introduced into the 3' genomic region of *Bmp5*, reporter gene expression was disturbed in specific parts of skeletal structures rather than in all the skeletal structures. Finally, those regulatory sequences identified in the 3'region of *Bmp5* were far apart from the *Bmp5* promoter. In fact, the distance between the lung mesenchyme regulatory sequence and the *Bmp5* transcription initiation site was more than 270kb (DiLeone, et al., 1998, 2000; Chandler et al., 2007).

We hypothesize that this 40Mb-QTL contains gene(s) along with possible regulatory elements that are responsible for the length and width phenotype in the femurs of Portuguese water dogs. In order to replicate the genetic mechanism of this particular bone phenotype and identify the gene(s) and possible regulatory element(s), we attempt to engineer the way to transfer the whole QTL into mouse embryonic stem cell by combining modern molecular techniques such as gene targeting, iPS cell technology and nuclear cell fusion technology. Generation of QTL phenotypes seemingly requires involvement of multiple genes. QTL of our interest, which spans in 40Mbs, may possibly harbor a group of genes generating this skeletal phenotype. The best chance to generate the phenotype and identify the responsible gene(s) of interest efficiently may be to move a whole fragment in the environment where we are able to manipulate the fragment in the way we want. Utilizing mice and their genetics enables us to do so because of advanced understandings of their genetics and available molecular tools. Transferring the QTL as a whole into mouse genome has another benefit, which is that all the canine transcription units and regulatory elements that are necessary for phenotypic

expression will be intact so that it will maximize the chance of generating the phenotype. It has been reported that transcript levels of mouse *Prx1* increased in developing forelimb bones and significant elongations of forelimbs were observed when a limb-specific enhancer of mouse Prx1 was replaced by the orthologous sequence from a bat (Cretekos et al., 2008). Therefore, even if an enhancer or repressor is solely responsible for the strong phenotype in Portuguese water dogs, this canine skeletal phenotype in mice by the canine enhancer/repressor will possibly be expressed and identified.

This project faces many major obstacles that have to be overcome. First and foremost, there is neither canine cell line nor ES cell with this particular QTL. In order to achieve the recombination between mouse and dog chromosome and create the mouse-dog chromosome with this specific QTL in mouse ES cells, the canine chromosome 12 with this QTL must be altered by gene targeting the modifying Cre-Lox cassette of our choice to canine chromosome 12. It has been reported that efficiency of gene targeting in somatic cells is significantly reduced compared to ES cells because of cell senescence, lower overall rates of homologous recombination and higher rate of nonhomologous end joining (Brown et al., 1997; Arbones et al., 1994; Hanson & Sedivy, 1995; Waldman, 1992). Therefore, it is necessary to have the canine ES cells that contain the QTL of interest in order to maximize the efficiency of gene targeting and the chance to make this project work. This requires us to generate canine ES cells from the fibroblasts harboring the QTL. As introduced in Chapter 1, after the exhausting search for key factors to make ES cells out of fully differentiated cells among numerous factors involving in maintenance of stem cell pluripotency, Yamanaka and colleagues (Takhashi & Yamanaka, 2006) have found that combination of the four transcription factors, *Oct4*,

Sox2, *Klf4* and *c-Myc*, is capable of turning somatic cells into the cells possessing pluriopotency when these factors are introduced into cells by retroviral virus. Since these four factors can induce pluriopotency of a stem cell in any somatic cell, the technology is named iPS cell technology, which has been utilized in this project. In this chapter I will focus on only on generation of canine ES cells out of the QTL-harboring fibroblasts and gene targeting in the canine immortalizing QTL-harboring fibroblast cells.

Materials and methods

Construction of the canine chromosome translocation vector

To isolate the homology arms of the targeting vector, canine lambda DNA library was constructed by using Lambda FIX II kit (Strategene) as following the instruction of the company's manual. The genomic DNAs were isolated from the canine fibroblasts that carry both extreme phenotypes (short/wide leg and long/narrow leg). The homology arms were identified by a canine DNA probe (F: catggcaacaaaatagcgaataaacgtttcccccagggcctgtctcagac; R:ggatttggtatcagttcaagataacctcggggcatgcaaacaccatggat). The isolated homology arms were cloned into the recombineering vector (pStart-K-Dog) for further cloning process. The chromosome transfer cassette (CAG-Cre-IRES-Puromycin-5'half-Neo) was cloned into the pStart-K-Dog using engineered AscI-Pme-I sites. After the completion of insertion of the chromosome transfer cassette, the vector was shuttled into the vector, pWS-TK by recombineering cloning. pWS-TK contained the negative selection, TK, which flanked the homology arms with the chromosome transfer cassette.

Construction of immortalizing *Piggybac* vector

Construction of the immortalizing *piggyBac* vector was described as following. hTERT was excised by NheI and SalI from pCI-Neo-hTERT (a gift from Dr. Robert Weinberg, MIT) and subcloned into engineered SpeI and XhoI of the shuttling vector, which has BglII and NheI sites engineered outside SpeI and XhoI. SV40 large T antigen cDNA was excised from pSG5 Large T (Addgene; Plasmid 9053) by EcoRI and BglII and subcloned into EcoRI and BamHI of the shuttling vector engineered ClaI and SpeI outside EcoRI and BglII. hTERT and SV40 large T antigen were excised by BglII/NheI and ClaI/SpeI and cloned into the CAGG promoter and EF1α of pVIVO4 (Invivogen), respectively. Finally, EF1α-hTERT-CAGG-SV40 large T antigen was cloned into pStart-K engineered with AscI by recombination-mediated genetic engineering, excised and cloned into *Piggybac* vector (pSP72- *Piggybac*MiniM1) by using AscI.

Preparation of canine primary fibroblast cells

Tissue samples were taken from four adult Portuguese water dogs that carry two extreme ends of the phenotypes (i.e., two with shortest and widest femurs and two with longest and narrowest femurs). Tissue samples were also collected from 3-month-old basset hounds that carry this particular QTL by ear punch. For isolation of adult canine fibroblasts, the tissue samples were incubated with 1X trypsin/EDTA for 15 minutes at 37°C and pipetted up and down a few times to dissociate the tissue during the incubation. After the inactivation of Trypsin/EDTA by fibroblast medium, the dissociated fibroblasts were plated and incubated at 37°C with 5% CO2 for further growth. For the isolation of the neonatal canine fibroblasts, the tissue samples were incubated with 0.1% Collagenase

at 37°C overnight with gentle rocking. After adding the fibroblast medium and settling in room temperature for a few minutes, the supernatant was collected and spun down for 5 minutes at 1000rpm. The dissociated fibroblast were re-suspended in the fresh fibroblast medium, plated and incubated at 37°C with 5% CO2 for further growth.

Lentiviral and *Piggybac* constructs

All mono-cistronic lentiviral stem cell vectors, FUW-Oct4, Sox2, Klf4, c-Myc , were obtained from Dr. Rudolf Jaenisch in MIT. A single polycistronic lentiviral stem cell vector (pHAGE-EF1α-STEMCCA) was obtained from Dr. Gustavo Mostoslavsky at Boston University. Lentiviral packaging vectors (Rev, Gag/Pol, VSV-G) for FUW-lentiviral stem cell vectors and GFP-expressing lentiviral vector (CSII-EF-MCS-IRES-GFP) were a generous gift from Dr.Eiji Morita at the University of Utah. Lentiviral packaging vectors (pHDM-Hgpm2, tat1b, G-corrected, and pRC/CMV-rev1b) for pHAGE-EF1α-STEMCCA were obtained from Dr. Jeng-Shin Lee at Harvard Medical School. *Piggybac* stem cell vectors carrying either human or mouse Oct4, Sox2, Klf4, c-Myc and Transposase-carrying vector (pKan-CAG-PBase) were generous gifts from Dr. Sen Wu at the University of Utah.

Lentiviral production and infection

The lentivirus was produced in HEK293T cells by calcium phosphate co-transfection of lentiviral shuttle vectors with three viral packaging vectors (Rev, Gag/Pol, VSV-G). For production of lentivirus by a single polycistronic lentiviral shuttle vector, additional viral Tat-expressing packaging vector was co-transfected with the single

polycistronic lentiviral shuttle vector. The viral production by calcium phosphate co-transfection in HEK293T cells was previously described (Miyoshi et al., 1998). Briefly, after co-transfetction of the viral vectors in HEK293T cells overnight (16-24 hour), the lentivirus was transduced in the fresh fibroblast medium (7.5ml) with 10μM Forskolin for 48 hours at 37°C with 5% CO_2. After the lentivirus was transduced for 48 hours, the virus-containing medium was collected, filtered through 0.2-μm pore size cellulose acetate filter and concentrated by ultracentrifugation at 16,500 rpm at 4°C if necessary. The virus was collected every 24 hours starting 48 hours after the virus was transduced. Eight hundred thousand canine fibroblasts, which were seeded in 100-mm culture plates 24 hours before the infection, were infected with filtered-concentrated virus in the presence of polybrene (5μg/ml). The fresh virus/polybrene-containing medium was changed every 24 hours and the canine fibroblasts were infection every 4 days. After 4 days infection, the fibroblasts were washed with PBS three times and incubated in the fresh fibroblast medium for 48 hours with an exchange of the fresh medium in 24 hours. Then, the infected fibroblasts were trypsinized with 0.05% trypsin and re-plated onto either the MEF feeder cells or gelatin-coated plates with canine ES medium. The canine ES medium was changed every 3 days. The infected canine fibroblasts were watched for their morphological change to stem cell-like morphology over 24 days after re-plated.

Cell culture

Canine iPS cells were cultured either on mitotically inactivated mouse embryonic (MEF) feeders or gelatin-coated plates in canine ES medium. Canine ES medium consists of Dulbecco's modified Eagle's medium (DMEM)/Ham's F-12 medium

(Invitrogen); 15% fetal bovine serum (FBS) (HyClone); 0.1mM nonessential amino acids (Invitrogen); 1mM L-glutamine, 1mM sodium pyruvate (Invitrogen); 1U/ml, 1μg/ml penicillin/streptomycin (Invitrogen); human leukemia inhibitory factor (hLIF) (10ng/ml, Millipore). Canine and murine fibroblast cells were always maintained in fibroblast medium, which consists of DMEM; 10% FBS; 0.1mM nonessential amino acids; 1mM L-glutamine, 1mM sodium pyruvate; 1U/ml, 1μg/ml penicillin/streptomycin.

Southern blot

Standard methods of Southern blot analysis were used to determine whether the chromosome translocation vector was incorporated into the beginning of QTL of interest. Southern blots were performed on DNA digested with NdeI for 5' end and SpeI fro 3' end of targeting site that cut once in genomic and the targeting vector internally and externally in order to determine whether the targeting vector was intact upon its insertion. When 5' end of the targeted allele is intact, a 12Kb band is expected for the nontargeted allele whereas a 8kb band is expected for the targeted allele. When 3' end of the targeting allele is intact, a 10Kb band was expected for the nontargeted allele whereas a 14Kb band is expected for the targeted allele Neomycin was re-probed on the Southern blot was previously performed for targeting detection to confirm transfection of the targeting vectors into cells.

Transfection of *Piggybac* stem cell and immortalizing vectors

Piggybac hTERT-SV40 immortalizing vectors with transposon containing plasmids were electroporated into $1x10^6$ adult canine fibroblasts, which were trypsined

by 0.05% trypsin, were collected into 1ml of electroporation buffer in electroporation cubet and placed on ice until the time of electroporation. The cells were placed and re-suspended into the fibroblast culture medium immediately after the electroporation was complete. Then, the cells were plated into 10cm culture plates and incubated at 37°C with 5% CO_2. The hygromycin selection was started 24 hours after the transfection and continued over 7-10 days before neomycin positive colonies were picked. For the *piggybac* stem cell factor vector, the electroporation procedure and selection procedure of identifying the positive clones for stem cell factor vectors in the canine immortalized cells were the same as described above except Neomycin selection was used to select for the stem cell factor positive dog immortalized cells.

Results

Construction of the canine chromosome translocation vector (CCTV)

It is essential to construct a targeting vector with the isogenic homology arm in order to maximize the targeting efficiency. Canine lambda genomic libraries were constructed from two Portuguese water dogs with the extreme phenotype (i.e. the dog with the longest and narrowest femurs or the one with the shortest and widest femurs) to isolate the homology arms. The results showed that the radiolabeled probe for the targeting gene, *GLUD1* identified the correct clones (Figure 2-1 Top panels). The identified clones were validated by the DNA sequencing.

Gene targeting of the CCTV into the canine immortalized fibroblasts

Ten million of the canine immortalized adult fibroblasts were used for gene targeting of the canine chromosome transfer vector. One hundred microgram of the targeting vectors were electroporated into those canine immortalized fibroblasts. When the targeting vector was integrated into the targeted site (GLUD1) in the canine chromosome 13, a 4-kb downshift (14-kb for WT to 8-Kb for the mutant) would be observed on the southern blotting that probed by the 5' canine DNA probe (Figure 2.1 bottom panel). A 4-Kb upshift should be observed on the southern blotting that was probed by the 3'canine DNA probe (Figure 2.1 bottom panel). The results showed that the 5'southern blotting showed no targeted fibroblast was recovered out of the 500 clones (Figure 2.2 left panel). Thus, the targeting of the chromosome transfer vectors did not occurred into the canine immortalized fibroblasts. To check if the targeting vectors were actually introduced into these host cells by the electroporation, the southern blotting that was probed for the neomycin-resistance gene was performed. The results of southern blotting showed the presence of the neomycin in those negative clones of the canine immortalized fibroblasts, which indicated that the electroporation worked properly to introduce the targeting vectors in the host cells (Figure 2.2 right panel).

Reprogramming of the adult canine fibroblasts by iPS technology

Gene targeting in the stem cells has been shown to be more effective than in the somatic cells. Canine embryonic stem cells with this specific QTL are not available. Recent developments in the regenerative medicine enable us to generate the embryonic

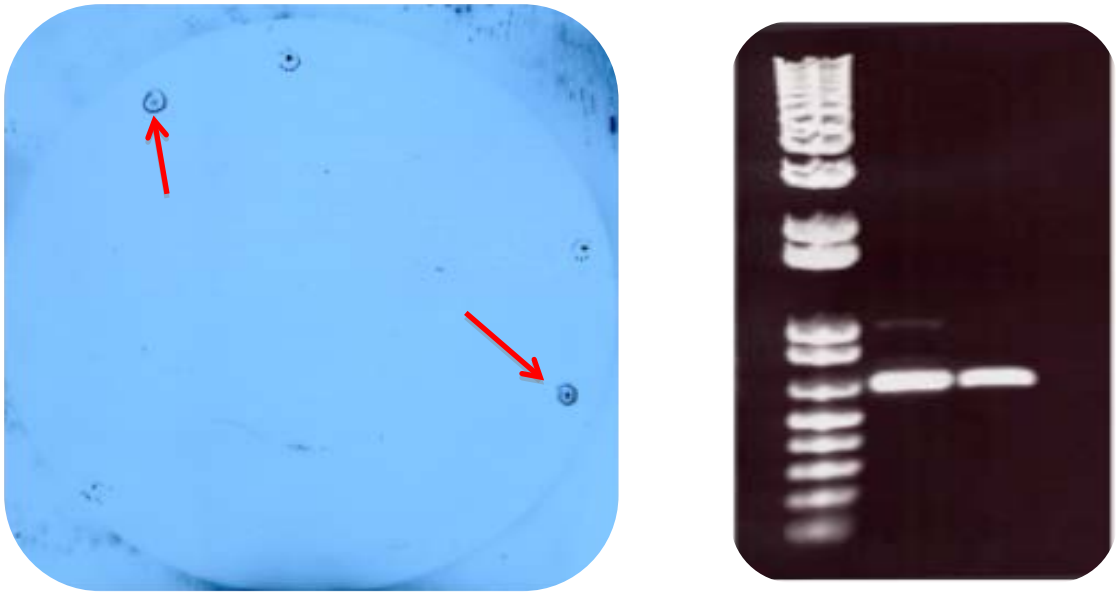

Figure 2- 1. Lambda DNA library screening of the canine DNA homology arms. The red arrows indicate the positive clones that identified by the canine GLUD1 probe (Top right). Positive identification of the isolated lambda clones that contain the canine GLUD1 homology arms (Top left). Southern Strategy to identify the targeted canine fibroblast clones. The blue lines indicate the 5'end southern blot whereas the black lines show the 3' end southern blot (Bottom panel).

stem cells out of the somatic cells by virally introducing four transcription factors, *Oct4, Sox2*, *Klf4*, and *c-Myc*, into the somatic cells. To explore this possibility, the adult canine primary fibroblasts were transfected with the lentivural vectors expressing these four reprogramming factors. Fourteen days after the viral infection of these four factors, the transfected adult canine primary fibroblasts started to form the stem cell-like colonies on the mouse feeder layers (Figure 2-3 top panels). These colonies indeed looked like

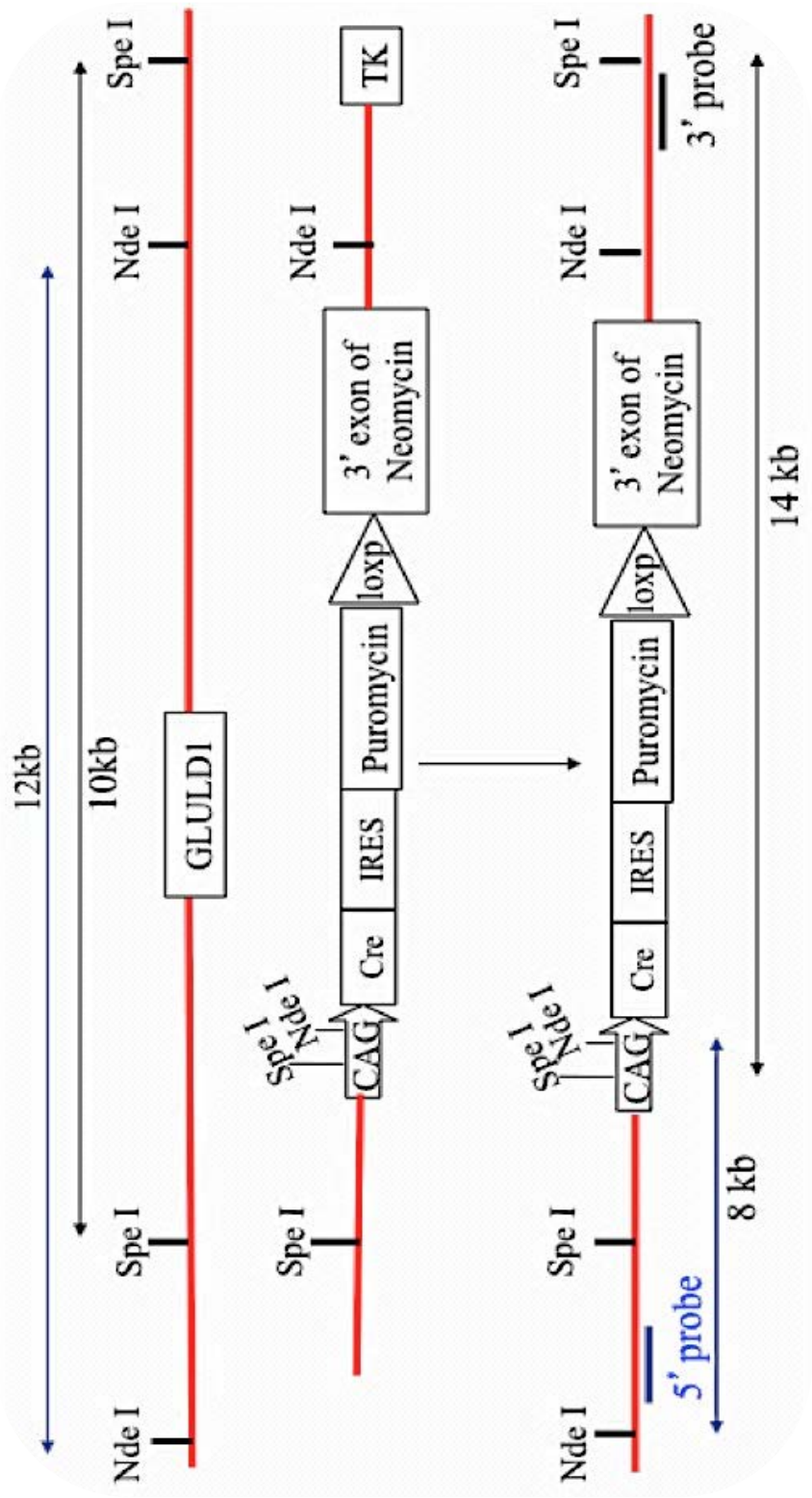

Figure 2-1 Cont.

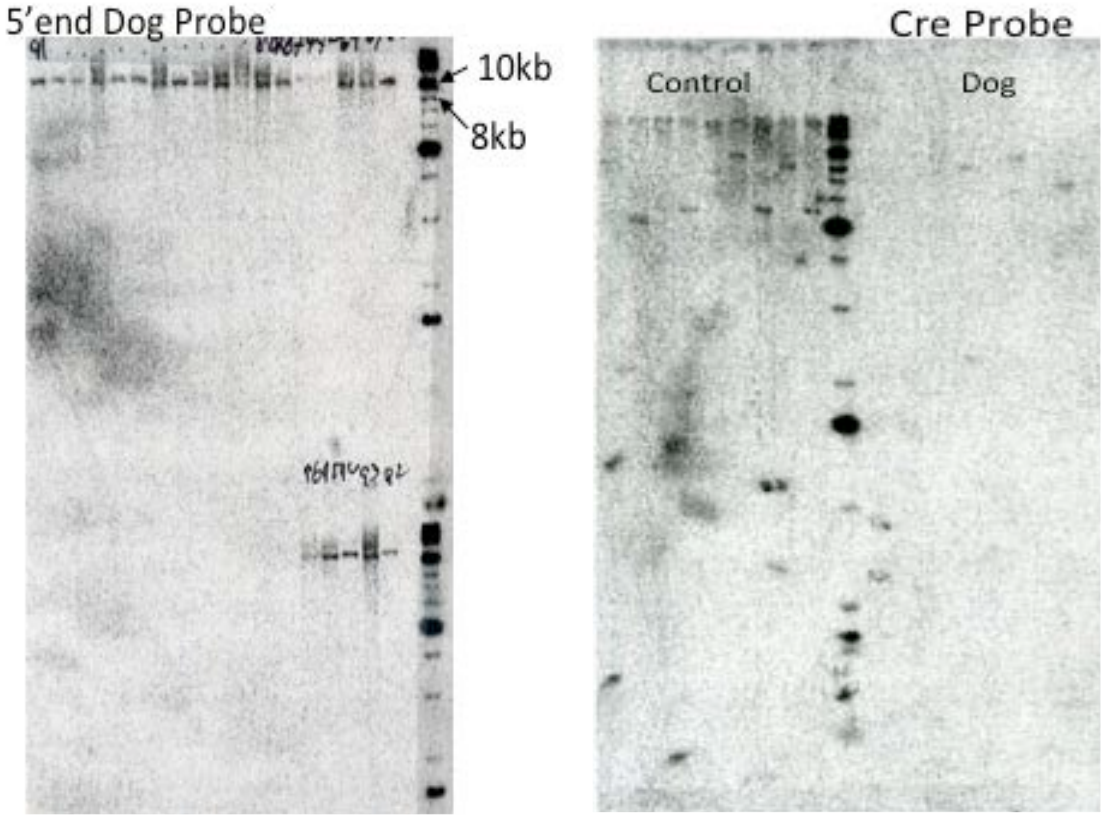

Figure 2- 2. Targeting to the canine GLUD1 in the canine chromosome 12. Southern blotting by the 5'canine DNA probe to detect the targeted integration of the targeting vectors (Left panel). Neomycin detection in the canine immortalized fibroblasts (Right panel).

the mouse ES cell colonies (Figure 2-3 bottom panels). However, after a few passages these canine iPSCs had stopped growing.

As described in Chapter 1, cell senescence is one of the major "roadblocks" to reprogram to pluripotency. To overcome this reprogramming roadblock, the immortalized cells are often used to generate iPSCs. Thus, the canine neonatal fibroblasts were isolated from the 3 month-old Beagles, which carried the QTL of interest. These neonatal fibroblasts as well as the canine adult fibroblasts were immortalized by the *piggyBac*-hTERT-SV40 vector. Although these canine immortalized fibroblasts emerged as the stem cell-like colonies, these canine stem cell-like colonies appeared to be the intermediate iPSCs. These colonies were never fully grown as observed earlier. Interestingly, the canine iPSC population may be heterogeneous because the two distinct appearances of the canine (intermediate) iPSCs were observed in culture. One looked like the mouse stem cells (Figure 2-3 top panels), whereas the other population looked like the canine ES cells, which was recently identified and isolated from one of the pure canine breeds, the beagle (Figure 2-4 top panels).

Discussion

An overall goal of this project was to develop a new genetic technology that allowed us to transfer a large fragment of a chromosome between different species. This genetic technology is the combination of a well-established technique, gene targeting with the Cre-Lox system, and a newly developing technique that enables the chromosome to cross the cell membrane. The canine QTL that inversely regulated the growth of

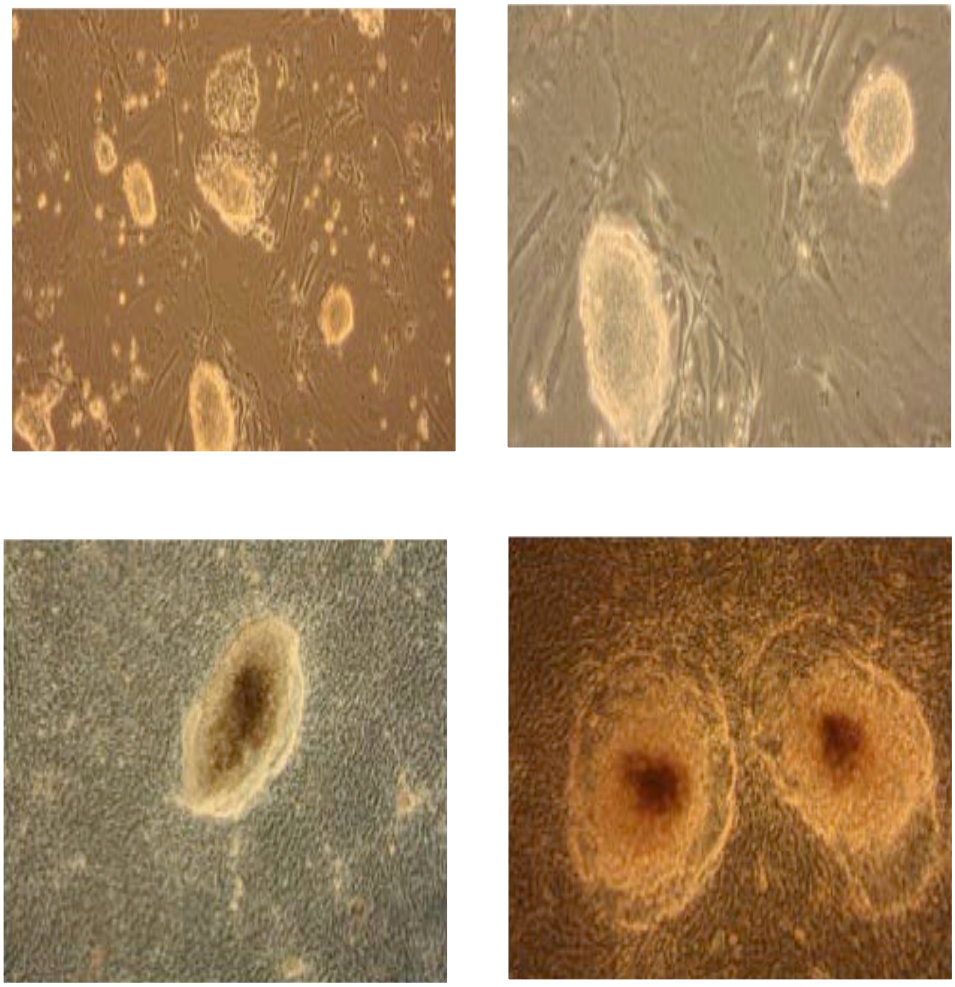

Figure 2- 3. Generation of the canine iPS cells from the canine primary adult fibroblasts (Top panels). The mouse embryonic fibroblasts were used to generate mouse iPS cells as positive controls (Bottom panels).

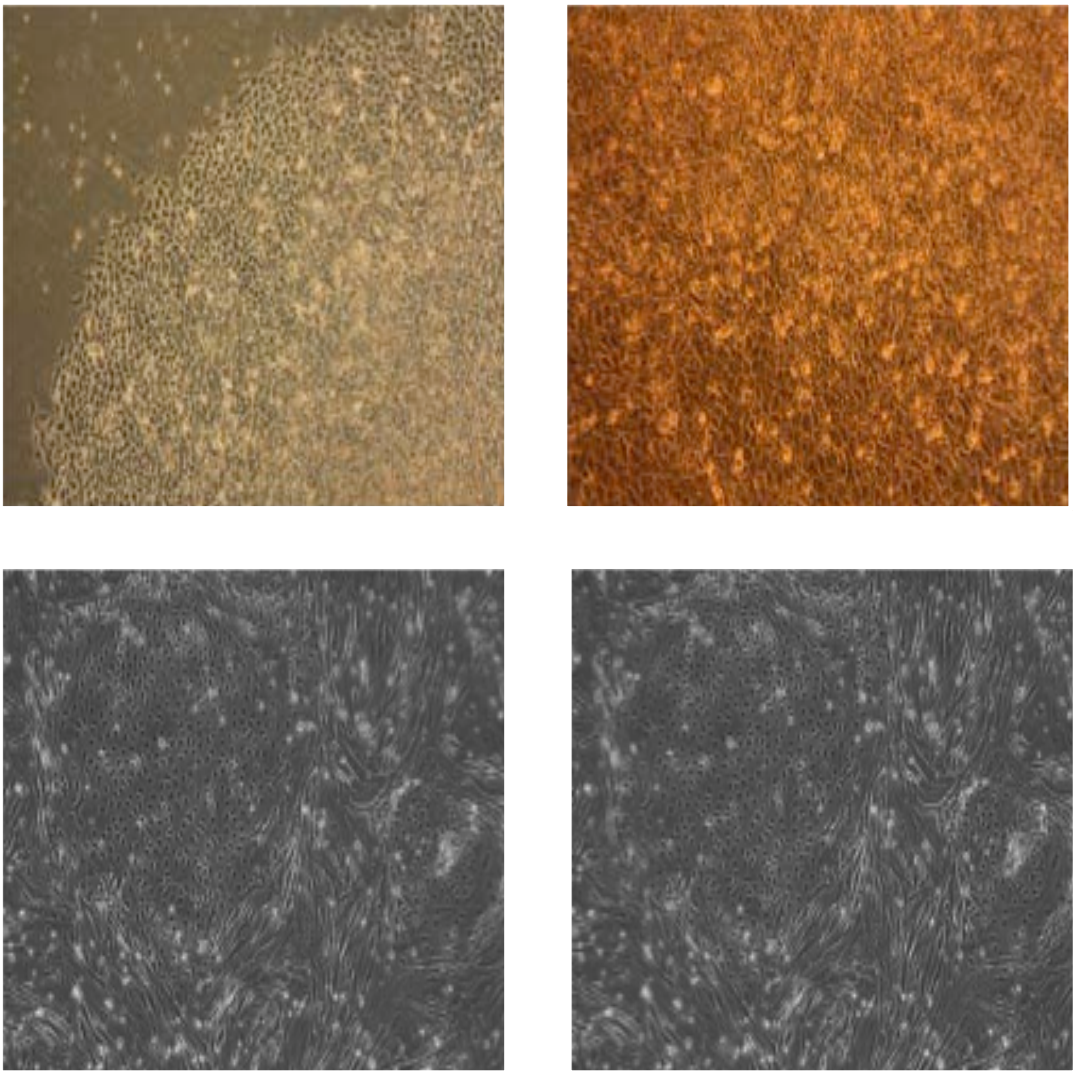

Figure 2- 4. iPSC formation from the canine immortalized neonatal fibroblasts by using the piggyBac reprogramming vectors. The canine stem cell-like colonies were found during the reprogramming process (Top panels). The images of the actual canine ES cells isolated from the Beagle (Bottom panels).

length and width was used as a model. This chapter mainly describes the challenges in application of gene targeting in the canine cells.

Gene targeting has been used in different cell types from different species (Mir and Piedrahita, 2004; Meehan et al., 2008). In this project gene targeting was attempted in the adult canine fibroblasts. Targeting efficiency is extremely low in primary somatic cells because of cell senescence, lower overall rates of homologous recombination and higher rate of nonhomologous end joining (Brown et al., 1997; Arbones et al., 1994; Hanson & Sedivy, 1995; Waldman, 1992). To overcome these problems and succeed in gene targeting, the canine adult fibroblasts were immortalized by coexpression of hTERT with SV40 largeT antigen because the stable transfection of hTERT alone was not sufficient to immortalize the primary canine fibroblasts. It has been reported that co-expression of hTERT with SV40 LargeT antigen was capable of immortalizing the primary cells that were difficult to transform without damage to the genome, such as genomic shuffling (O'Hare et al., 2001).

As the results showed, the chromosome transfer vectors were indeed introduced into the canine immortalized fibroblasts, but the genomic incorporation of the vectors was always random. Although gene targeting can be successfully performed in immortalized somatic cells from different species, the success of gene targeting in the immortalized somatic cells requires modifications, including cell-cycle synchronization and the addition of the nuclear localization sequences (Mir and Piedrahita, 2004; Meehan et al., 2008). Even with the modifications, the efficiency of gene targeting is extremely low in those somatic cells compared to the targeting efficiency in mouse embryonic stem

cells (Mir and Piedrahita, 2004; Meehan et al., 2008). For success of this entire project, it is essential to obtain the canine embryonic stem cells with the QTL of our interest.

Recent invention of iPSC technology has enabled researchers to generate stem cell-like cells from the somatic cells of different species such as mice, rats, and humans (Takahashi and Yamanaka, 2006; Liao et al., 2009; Takahashi et al., 2007). iPSC technology was tried in the adult canine fibroblasts to see if canine iPSCs could be generated. The initiation of canine iPSC generation was observed yet those iPS cells stopped growing after they were isolated by using trypsin, which reportedly harmed the canine embryonic stem cells during their isolation (Vaags et al., 2009; Hayes et al. 2008). Although an alternative enzyme dispase was used for the isolation of the canine iPSC colonies later on, the canine adult fibroblasts could be partially reprogrammed but could not complete the reprogramming process.

The process of iPSC generation has two phases: iPSC generation phase and maintenance phase. In each phase, specific molecular and genetic events must occur in the cells in order for the cells to obtain the full pluropotency as fully competent iPSCs (Feng et al., 2009; Jaensch and Young, 2008). It has been identified that the number of small molecule inhibitors are capable of promoting each phase of iPSC generation (Feng et al., 2009). A triple cocktail of valproic acid, sodium butyrate and vitamin C (termed as 3i) has worked the best to promote each phase (personal communication with Dr. Wu, University of Utah). Even with the supplement of the 3i, the canine adult fibroblasts never become the fully competent iPSCs.

Both the mouse and the human iPSC four factors were tried to generate the canine iPSCs. However, neither the mouse nor the human factors could complete the

reprogramming processes in the canine primary fibroblast or the immortalized fibroblasts. Possibly, neither the mouse nor the human reprogramming factors were compatible to the canine transcriptional machinery. Shimada and colleagues recently demonstrated the first success of the canine iPSC generation (Shimada, et al., 2010). Importantly, Shimada and colleagues used canine *Oct4, Sox2, Klf4,* and *c-Myc* with the small molecule inhibitor, valproic acid to generate the canine embryonic fibroblasts (Shimada et al., 2010). Canine iPSCs could be generated from the canine adult fibroblasts if these fibroblasts were reprogrammed by the canine *Oct4, Sox2, Klf4,* and *c-Myc* with the supplement of 3i (valproic acid, sodium butyrate and vitamin C). In sum, both applications of gene targeting and iPSC techonology to canine immortalized fibroblasts were unsuccessful because of the possible reasons described above. However, with technical improvements described above, both gene targeting and canine iPSC formation will be achievable in the future.

References

Adams, S.L., Cohen, A.J. and Lassova, L. (2007). Integration of signaling pathways regulating chondrocyte differentiation during endochondral bone formation. J Cell Physiol 213, 635-641.

Arbones, M.L., Austin, H.A., Capon, D.J. and Greenburg, G. (1994). Gene targeting in normal somatic cells: inactivation of the interferon-gamma receptor in myoblasts. Nature Genet., 6, 90-97.

Bondarenko, V.A., Liu, Y.V., Jiang, Y.I. and Studitsy, V.M. (2003). Communication over a large distance: enhancers and insulators. Biochem Cell Biol 81, 241-251.

Brown, J.P., Wei, W. and Sedivy, J.M. (1997). Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. Science, 277, 831-834.

Brunet, L.J., McMahon, J.A., McMahon, A.P. and Harland, R.M. (1998). Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. Science 280, 1455-1458.

Carrier, D.R., Chase, K. and Lark, K.G. (2005). Genetics of canid skeletal variation: Size and shape of the pelvis. Genome Res 15, 1825-1830.

Chase, K., Adler, F.R., Miller-Stebbings, K. and Lark, K.G. (1999) Teaching a new dog old tricks: identifying quantitative trait loci using lessons from plants. J Hered 90, 43-51.

Chase, K., Carrier, D.R., Adler, F.R., Jarvik, T., Ostrander, E.A., Lorentzen, T.D. and Lark, K.G. (2002). Genetic basis for systems of skeletal quantitative traits: Principle component analysis of the canid skelton. Proc Natl Acad Sci USA 99, 9930-9935.

Chandler, R.L., Chandler, K.J., McFarland, K.A. and Mortlock, D.P. (2007). Bmp2 transcription in osteoblast progenitors is regulated by a distant 3' enhancer located 156.3kilobases from the promoter. Mol Cell Biol 27, 2934-2951.

Counter, C.M., Meyerson, M., Eaton, E.N., Ellisen, L.W., Caddle, S.D., Haber, D.A. and Weinberg, R.A. (1998). Telomeease activity is resorted in human cells by ectopoic expression of hTERT (hEST2), the catalytic subunit of telomerase. Oncogene 16, 1217-1222.

Cretekos, C.J., Wang, Y., Green, E.D., Martin, J.F., RasweilerIV, J.J. and Behringer, R.R. (2008). Regulatory divergence modifies limb length between mammals. Genes Dev 22, 141-151.

Dekker J. (2008). Gene regulation in third dimension. Science 319, 1793-1794.

DiLeone, R.J., Russell, L.B. and Kingsley, D.M. (1998). An extensive 3' regulatory region controls expression of Bmp5 in specific anatomical structures of the mouse embryo. Genetics 148, 401-408.

DiLeone, R.J., Marcus, G.A., Johnson, M.D. and Kingsley, D.M. (2000). Efficient studies of long-distance Bmp5 gene regulation usig bacterial artificial chromosomes. Proc Natl Acad Sci USA 97, 1612-1617.

Engel, J.D. and Tanimoto, K. (2000). Looping, linking and chromatin activity: New insights into β-globin locus regulation. Cell 100, 499-502.

Feng, B., Ng, J-H., Heng, J-C., D. and Ng, H-H. (2009). Molecules that promote or enhance reprograming of somatic cells to induced pluripotent stem cells. Cell Stem Cell, 4, 301-312.

Hanson, K.D. and Sedivy, J.M. (1995). Analysis of biological selections for high-efficiency gene targeting. Mol. Cell. Biol., 15, 45-51.

Hartl, D.L. (2000). A primer of population genetics. Sinauer Associates, MA.

Hayes, B., Fagerlle, S.R., Ramarkrishnan A., Baran, S., Harkey, M., Graf, L., Bar, M., Bendoralte, A., Tewari, M. and Torok-Storb, B. (2008). Derivation, characterization, and in vitro differentiation of canine embryonic stem cells. Stem Cells. 26, 465-473.

Huang, J.D., Schwyter, D.H., Shirokawa, J.M. and Courey, A.J. (1993). The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. Genes Dev 7,694-704.

Inada, M., Yasui, T., Namura, S., Miyake, S., Deguchi, K., Himeno, M., Sato, M., Yamaguchi, H., Kimura, T., Yasui, N., Ochi, T., Endo, N., Kitamura, Y., Kishimoto, T. and Komori, T. (1999). Maturational disturbance of chondrocytes in Cbfa1-deficient mice. Dev Dyn 214, 279-290.

Jackson, J.E. (2001). A user's guide to principle components. Wiley-interscience, New York.

Jesen, E.D., Schroeder, T.M., Bailey, J., Gopalakrishnan, R. and Westendorf, J.J. (2008). Histone deacetylase 7 associates with Runx2 and rrepresses its activity during osteoblast maturation in a deacetylation-independent manner. J Bone Miner Res 23, 361-372.

Kadauke, S. and Blobel, G.A. (2008). Chromatin loops in gene regulation. Biochim Biophys Acta.

Karsenty, G. (2003). The complexities of skeletal biology. Nature 423, 316-318.

Kaufmann, E. and Knochel, W. (1996). Five years on the wings of fork head. Mech Dev 57, 3-20.

Kawamura, N., Kugimiya, F., Oshima, Y., Ohba, S., Ikeda, T., Saito, T., Shinoda, Y., Kawasaki, Y., Ogata, N., Hoshi, K., Akiyama, T., Chen, W.S., Hay, N., Tobe, K., Kadowaki, T., Azuma, Y., Tanaka, S., Nakamura, K., Chung, U.I. and Kawaguchi, H. (2007). Akti in osteoblasts and osteoclasts controls bone remodeling. PLoS ONE 2, e1058.

King, J.A., Marker, P.C., Seung, K.J. and Kingsley, D.M. (1994). BMP5 and the molecular, skeletal and soft-tissue alterations in short ear mice. Dev Biol 166, 112-122.

Kingsley, D.M. (1994). The TGF-β family: new members, new reporters, and new genetic tests of function in different organisms. Genes Dev 8, 133-146.

Kingsley, D.M., Bland, A.E., Grubber, J.M., Marker, P.C., Russell, L.B., Copeland, N.G. and Jenkins, N.A. (1994). The mouse short ear skeletal morphogenesis locus is

associated with defects in a bone morphogenetic member of the TGF beta superfamily. Cell 71, 399-410 .

Kleinjan, D.A. and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76, 8-32.

Kobayashi, T. and Konenberg, H. (2005). Minireview: transcriptional regulation in development of bone. Endocrinology 146, 1012-1017.

Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R.T., Gao, Y.H., Inada, M., Sato, M., Okamoto, R., Kitamura, Y., Yoshiki, S. and Kishimoto, T. (1997). Targeted disruption o Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell 30, 755-764.

Kronenberg, H.M. (2003). Developmental regulation of the growth plate. Nature 423, 332-336.

Lark, K.G., Chase, K. and Sutter, N.B. (2006). Genetic architecture of the dog: sexual size dimorphism and functional morphology. Trends Genetics 224, 537-544 .

Lee, G.R., Kim, S.T., Spilianakis, C.G., Fields, P.E. and Flavell, R.A. (2006). T helper cell differentiation: regulation by cis elements and epigenetics. Immunity 24, 369-379.

Lehmann, O.J., Sowden, J.C., Calsso, P., Jordan, T. and Bhattacharya, S.S. (2003). Fox's in development and disease. Trends Genetics 19, 339-344.

Lettice L. (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci USA 99,7548-7553.

Lettice, L., Heaney, S., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E. and de Graff, E. (2009). A long-range shh enhancer regulates expression in the developing limb and fin and is associated with prreaxial polydactyly. Hum Mol Genet 12, 1725-1735.

Li, G., Barkess, G. and Qian, H. (2006). Chromatin looping and the probability of transcription. Trends Genet 22, 197-202.
Liao, J., Cui, C., Chen, S., Ren, J., Chen, J., Gao, Y., Li, H., Jia, N., Chend, L. and Xiao, H. (2009). Generation of induced pluripotyent stem cell lines from adult rat cells. Cell Stem Cell, 4, 11-15.

Limour, V., Dentry, C., Sanchez, C., Henrontin, Y., Castronovo, V. and Bellahcene, A. (2007). Runx2- and histone deacetylase 3- mediated repression is relieved in differentiating human osteoblast cells to allow high bone sialoprotein expression. J Bio Chem 282, 36240-362490.

Manak, J.R., Mathies, L.D. and Scott, M.P. (1994). Regulation of decapentaplegic midgut enhancer by homeotic proteins. Development 120, 3605-3619.

Miele, A. and Dekker, J. (2008). Long-range chromosomal interactions and gene regulation. Mol Biosyst 4, 1046-1057.

Meehan, D. T., Zink, M. A., Mehlen, M., Nelson, M., Sanger, W. G., Mitalipov, S. M., Wolf, D. P., Ouellette, M.M. and Norgren Jr, R. B. (2008). Gene targeting in adult rhesus macaque fibroblasts. BMC Biotechnol., 8.

Minina, E., Wenzel, H.M., Kreschel, C., Karp, S., Gaffield, W., McMahon, A.P. and Vortkamp, A. (2001). BMP and Ihh/PTHrP signaling interact to coordinate chondrocytes proliferation and differentiation. Development 128, 4523-4534.

Minina, E., Kreschel, C., Naski, M.C., Ornitz, D.M. and Vorkamp, A. (2002). Interaction of FGF. Ihh/Pthih, and BMP signaling integrates chondrocytes proliferation and hypertrophic differentiation. Dev Cell 3, 439-449.

Mir, B. and Piedrahita, J. A. (2004). Nuclear localization signal and cell synchrony enhance gene targeting efficiency in primary fetal fibroblasts. Nucleic Acid Res., 32, e25.

Miura, T., Mattson, M.P. and Rao, M.S. (2004). Cellular lifespan and senescence signaling in embryonic stem cells. Aging Cell 3, 333-343.

Miyoshi, H., blomer, U., Takahashi, M., Gage, H. and Verma, I. (1998). Development of a self-inactivating lentivirus vector. J Virol., 72, 8150-8157.

O'Hare, M.J., Bond, J., Clarke, C., Takeuchi, Y., Atherton, A., Berry, C., Moody, J., Siliver A., Davies, D.C., Alsop, A.E., Neville, A.M. and Jat, P.S. (2001). Conditional immortalization of fresh isolated human mammary fibroblasts and endothelial cells. Proc Natl Acad Sci USA, 98, 646-651.

Ostrander, E.A., Giger, U. and Lindblad, K. (2005). The dog and its genome. Cold Spring Harbor Laboratory Press New York, US.

Ostrander ,E.A. and Wayne, R.K. (2005). The canine genome. Genome Res 15, 1706-1716.

Otto, F., Thornell, A.P., Cromptom, T., Denzel, A., Gilmour, K.C., Rosewell, I.R., Stamp, G.W., Beddington, R.S., Olsen, B.R., Selby, P.B. and Owen, M.J. (1997). Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell 30, 765-771.

Pizette, S. and Niswander, L. (2000). BMPs are required at two steps of limb chondrogensis: formation of precondrogenic condensation and their differentiation into chondrocytes. Dev Biol 209, 237-249.

Schroeder, T.M., Jensen, E.D. and Westendorf, J.J. (2005). Runx2: a master organizer of gene transcription in developing and maturing osteoblasts. Birth defects Res C Embryo Today. 75, 213-25.

Shimada, H., Nakada, A., Hashimoto, Y., Shigeno, K., Shionoya, Y. and Nakamura, T. (2010) Generation of canine induced pluripotent stem cells by retroviral transduction and chemical inhibitors. Mol. Reprod. Dev., 77, 1-2.

Spiliankis, C.G. and Flavell, R.A. (2004). Long-range intrachromsomal interaction in the T helper type2 cytokine locus. Nat Immunol 5, 1017-1027.

Stadler, H.S., Higgins, K.M. and Capecchi, M.R. (2001). Loss of Eph-receptor expression correlated with loss of cell adhesion and chondrogenic capacity in Hoxa13 mutant limbs. Development 128, 4177-88.

Stock, M. and Otto, F. (2005). Control of RUNX2 isoform expression: the role of promoters and enhancers. J Cell Biochem 95, 506-517.

Sutter, N.B. and Ostrander, E.A. (2004). Dog star rising: the canine genetic system. Nat Rev Genet 5, 900-910.

Takeda, S., Bonnamy, J.P., Owen, M.J., Ducy, P. and Karsenty, G. (2001). Continous expression of Cbfa1 in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues Cbfa-1deficeint mice. Genes Dev 15, 467-481.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 131, 861-872.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluriopotent stem cell from mouse embryonic and adult fibroblast cultures by defined factors. Cell, 25, 663-676.

Vaags, A.K., Rosic-Kablar, S., Gartley, C.J., Zheng, Y.Z., Chesney, A., Villgomez, D.A., Kruth, S.A. and Hough, M.R. (2009). Derivation and characterization of canine embryonic stem cell lines wit in vitro and in vivo differentiation potential. Stem Cells. 27, 329-340.

Vega, R.B., Matsuda, K., Oh, J., Barbosa, A.C., Yang, X., Meadows, E., McAnally, J., Pomajzl, C., Shelton, J.M., Richardson, J.A., Karsenty, G. and Olson, E.N. (2004). Histone deacetylase 4 controls chondrocyte hypertrophy during skeletogenesis. Cell 119, 555-56.

Waldman, A.S. (1992). Targeted homologous recombination in mammalian cells. Crit. Rev. Oncol. Hematol., 12, 49-64.

Wayne, R.K. and Ostrander, E.A. (2005). Lessons learned from the dog genome. Trends Genet 23, 557-567.

West, A.G. and Fraser, P. (2005). Remote control of gene transcription. Hum Mol Genet 14, R101-R111.

Wu, X., Shi, W. and Cao, X. (2007). Multiplicity of BMP signaling in skeletal development. Ann NY Acd Sci 1116, 2649.

Zuniga, A. (2005). Globalisation reaches gene regulation: the case for vertebrate limb development. Curr Opin Genet Dev 15, 403-409.

CHAPTER 3

CHARACTERIZATION OF A NOVEL GENE, AK045681 IN MOUSE BRAIN

Introduction

The mammalian central nervous system (CNS) is made up of diverse neuronal cell types, which is the reflection of the complexity of gene expression profiles that are present in the brain. The human genome project has revealed that the human genome consists of nearly 30,000 genes of which almost one third to one half are particularly expressed in the brain and for which the biological roles of the vast majority remain unknown (Diaz, 2009; van der Burg et al., 2010). A number of mutagenesis techniques are available for generating mutant mice, including chemical-induced mutagenesis, radiation-induced mutagenesis, and gene targeting. Each of these techniques is very powerful and each has its own advantages and disadvantages. For example, chemical mutagens such as ethyl-nitrosourea (ENU) produce random single base-pair changes throughout the genome with a high frequency of mutant phenotypes (Carlson & Largaespda, 2005). With this technique, however, researchers have no control over the introduction of the mutations. Also, this technique lacks useful molecular markers for

easy identification of the mutation responsible for the phenotypes. Gene targeting allows researchers great specificity for introducing the mutation of interest using molecular markers. As revolutionary and useful as this technology is, gene targeting can be technically challenging in terms of engineering the targeting vectors. Low targeting efficiency, frequent random integration of vectors, ES cell work and generations of mutant mice also make gene targeting more challenging and time-consuming (Carlson & Largaespda, 2005). To uncover biological functions of those unknown genes in the brain, high throughput mutagenesis approaches with high efficiency and site specificity of the vector integration will be required. Transposable elements or transposons are mobile genetic elements that move around via a "cut-and-paste" mechanism within the genome of an organism (Copeland & Jenkins, 2010; Izsvak & Ivics, 2004; Mates et al., 2007: Carlson & Largaespda, 2005). Because of their ability to insert themselves close to or within genes, transposons can be used to inactivate genes through insertional mutagenesis. This provides researchers with information about the function of the genes. Transposable elements have been modified and utilized as effective germ-line transgenesis and insertional mutagenesis in invertebrates such as flies and worms (Copeland & Jenkins, 2010; Izsvak & Ivics, 2004; Mates et al., 2007: Carlson & Largaespda, 2005). However, no functional transposon in vertebrate had been reported until the recent resurrection of the active Tc1-like DNA transposon, *Sleeping Beauty* (SB), from salmoind fishes by comparative phylogenic approach (Ivics, et al., 1997). Following the discovery of SB, a new DNA transposon, *piggybac* (PB) was discovered from the cabbage looper moth and proven to be functional in mammalian genomes such as mice and humans (Ding et. al, 2005). Both SB and PB do not integrate randomly into

target DNA and both show specificity of integration site such as TA dinucleotide for SB and TTAA tetranucleotide for PB (Copeland & Jenkins, 2010; Izsvak & Ivics, 2004; Mates et al., 2007: Carlson & Largaespda, 2005; Ding et. al., 2005). PB has great functional advantages over SB. It has been shown that upon excision of the transposable elements, SB leaves a 3-bp transposon "footprint" because of its excising mechanism (Copeland & Jenkins, 2010; Izsvak & Ivics, 2004; Mates et al., 2007: Carlson & Largaespda, 2005). The precise excision of the PB transposon mechanism, however, leaves no "footprint" behind (Ding et al., 2005; Thibault et al., 2004; Fraser et al., 1996; Elick et al., 1996). Furthermore, SB exhibits a strong tendency for local hopping. Local hopping is the phenomenon that the transposons preferentially reintegrate near its original insertion site upon excision of the transposons from the genome. In contrast, PB does not display any propensity for local hopping (Copeland & Jenkins, 2010; Izsvak & Ivics, 2004; Mates et al., 2007: Carlson & Largaespda, 2005; Ding et. al., 2005). PB is capable of carrying a much larger cargo (up to 20-kb) in its transposon vector than SB is (up to 10-kb) (Copeland & Jenkins, 2010; Izsvak & Ivics, 2004; Mates et al., 2007: Carlson & Largaespda, 2005; Ding et. al., 2005). Finally, it has been reported that *piggybac* can be encoded by transgenes and expressed via inducible or tissue-specific promoters to mobilize unlinked transposons in a regulated manner (Wu et al., 2007; Cadinanos & Bradley, 2007). Given the flexibility of molecular engineering, insertion specificity, and high transposition activity, *piggyback* shows great potential as a tool for mammalian transgenesis with the possibility of targeted integration.

Wu and colleagues have improved the gene-trap method with the *piggybac* transposon system. They identified the minimum length of the inverted terminal DNA

repeats requiring the highest transposition efficiency. They also elegantly designed their gene trap construct, which included β-gal reporter and constitutively expressed eGFP with a combination of Lox sites and FRT site. The β-gal expression reports an activity of the gene that the gene-trap cassette disrupts, and the constitutively expressed eGFP serves as a convenient marker for identification of new transposition in vivo. Most importantly, their versatile gene-trap construct can be conditionally excised by tissue-specific Cre expression so that the normal function will be "rescued" from the loss of the function state in the tissues of interest (Wu et al., 2007). Their *piggybac* gene trap screening in mice mutagenized an interesting novel gene, AK045681.

A novel gene locus, AK045681 is located in chromosome 12 in the mouse genome. It is made out of four exons. Exons 1 through 3 of this novel gene consist of a large 5 prime untranslated region (approximately 1100-bp), whereas exon 4 harbors the open reading frame, which is translated into 120 amino acids (AAs). Although the biological function of this small protein remains unknown, it has been speculated that AK045681 may play a role in RNA splicing, based on the AAs sequence similarity of AK045681 to the spliceosomal protein U1C. The spliceosomal U1C protein plays a critical role in the initiation and regulation of pre-mRNA splicing as a component of the U1small nuclear ribonucleoprotein particle (Patel & Steitz, 2003; Wahl et al., 2009). U1C consists of unusual proline and methionine-rich residues in the carboxy terminus and several cysteines and histdines in the amino terminus. These cysteine and histidine residues are arranged to form a zinc finger-like domain, which is known to be involved in protein-nucleic or protein-protein interactions (Nelissen et al., 1991; Forch et al., 2000, 2002; Muto, et al., 2004). AK04681 does not possess the zinc finger-like domain;

however, it does contain particular proline and methionine-rich sequences (PRS). The PRSs motifs are recognized by proline-rich sequence recognition domains (PRD) such as SH3, EVH1, GYF, UEV and WW domains (Kofler et al., 2009; Kofler & Freund, 2006). Complexes of PRS/ PRD are widely used by nature in higher eukaryotes and have participated in numerous biological processes including ubiquitin-mediated protein degradation, transcription, RNA processing, and cytoskeletal regulation (Kofler & Freund, 2006; Ingham et al., 2005). Frequently, the PRS motifs are found in splicing factors and spliceosomal proteins including the U1C, SmB/B', branch point-binding proteins and U5snRNP-asscociated protein (Kofler et al., 2004; Abovich and Rosbash, 1997; Wiesner et al., 2002). The PRS motifs provide docking sites for the PRDs (e.g., GYF and WW domains) and likely help the formation of the spliceosome. For example, the branch point-binding protein (BBP) in yeast and splicing factor 1 (SF1) in humans are RNA-binding proteins known to bind the branch point sequence during RNA splicing (Bedford et al., 1997). It has been shown that the carboxy terminus of BBP contains PRS, which is associated with Prp40 (pre-mRNA processing protein 40), a part of the U1snRNP in yeast (Abovich and Rosbash, 1997). The PRS of human SF1 is also known to bind to the WW domain of FBP1 (formin-binding protein 1), the vertebrate homologue of Prp40 (Bedford et al., 1997). The splicing factor SIPP1 (**S**plicing factor that **I**nteracts with **P**QBP-1 and P**P1**) has two PRS regions, which are able to interact with the WW domain of PQBP1 (polyglutamine tract-binding protein 1), and act as pre-mRNA splicing activators in intact cells (Llorian et al., 2004, 2005). Furthermore, it has been shown that SIPP1 also interacts with the WW domain of the human FBP21, and functions as a pre-mRNA splicing activator in vivo (Huang et al., 2009). As stated above, the presence of

PRS motifs in AK045681 and the similarity in AA sequences between the spliceosomal U1C protein and AK045681 have led me to hypothesize that pre-mRNA splicing has a potential biological role in AK045681.

This chapter describes the cellular and functional characterization of the novel gene, AK054681 in mice. The present study shows that AK045681 is postnatally expressed in GABAergic neurons in the specific neuronal regions (e.g., the amygdala and the hippocampus) that regulate anxiety-related behaviors. Neuronal nuclear co-localization of this novel gene to the spliceosomal SmB/B' protein supports the potential role of AK045681 in pre-mRNA splicing as a splicing factor, or a part of the spliceosome. Finally, disruption of AK045681 in mice causes interference of the expression of anxiety-related behaviors.

Materials and methods

Animals

Generation of AK045681 mutant mice was described elsewhere (Wu et al., 2007). A gene-trap vector, ZG-s consisted of IRES-β-galactosidase-pA and CAG-eGFP-pA cassette flanked by LoxP sites with the minimum LTRs. The minimum LTRs are required for the transposition. The entire cassette would be excised by mating with Cre-expressing mice for conditional rescue allele. The CAG-eGFP-pA cassette was flanked by FRT sites that can be excised by mating with Flpe-expressing mice at the researcher's convenience. Both ZG-s and *piggyBac* transposase-expressing construct were used for pronuclear injection to create the transgenic mouse lines. The insertion site of the gene-trap vectors was identified by inverse PCRs as described (Wu et al., 2007).

Cortical neuronal culture and DNA transfection

Mouse embryos (E13.5) were used to isolate the cortical neurons. Pregnant C57BL/6 female mice were sacrificed by cervical dislocation and the fetuses were transferred to sterile PBS. The cortices were dissected from the fetuses and single-cell suspension of the cortical cells was prepared by papain digestion followed by trituration through a P-1000 pipette and DNaseI treatment. The single-cell suspension of the cortical cells was washed by the dissection medium (Neurobasal medium with B12) and spun down at 1000rpm for 5 minutes twice. The isolated cortical cells were re-suspended in the neuronal culture medium (Neurobasal medium with B12 and bFGF) and were plated onto the poly-L-lysine coated culture dishes (Thermo Scientific).

The cortical neurons were transfected with either a plasmid carrying full-length cDNA of AK045681 or a spliced variant of AK045681cDNA. X-tremeGENE HP DNA transfection reagent (Roche) was used to transfect the aforementioned plasmids to the cortical neurons. The procedure was performed as the manufacture's instructed. Briefly, the DNA plasmids were diluted with serum-free medium to a final concentration of 1μg DNA per 100μl medium. The X-tremeGENE HP DNA transfection reagent (4:1 reagent to DNA ratio) was directly added to the medium with the diluted DNA plasmids. The transfection reagent/DNA complex was incubated for 15 minutes at room temperature before it was applied to the neurons in a dropwise manner. The cultured neurons were incubated with the transfection reagent/DNA complex for 48 hours before the further procedure.

Northern blotting

Total RNAs were isolated from the whole brains of AK045681 mutant mice and their wild type littermates (postnatal day 15 or older) with TRI Reagent (Molecular Research Center, Inc) by following the manufacture's instruction. Poly (A) mRNAs were further isolated from those total RNAs by using Dynabead-oligo (dT) (invirogen). Standard northern blotting was performed as described elsewhere (Sambrook and Russell, 2001). The probes for AK045681 exon3/4 were: exon3/4: F:ccaacctagcatgagcacgc; R:catttctctactagataaagagg; exon4-F:atgaagagcaggcccagaac;R:tgcatctcgtaaacactcag)

Reverse transcriptase PCR

Total RNAs were isolated from the whole brains of AK045681 mutant mice and their wild type littermates (postnatal day 15 or older) by TRI Reagent (Molecular Research Center, Inc) by following the manufacture's instructions. Any contaminating DNA was removed by DNseI treatment from the RNA preparation. Complementary DNA of AK045681 was synthesized with Superscript II reverse transcriptase 1st strand cDNA according to the manufactures' instruction. One microgram of RNA from each sample was used in each 20ul RT reaction, and 2ul of the cDNA products were used as template in each 20ul PCR reaction. The primer sequences for exon 3-4 transcript are: exon3/4-F:ccaacctagcatgagcacgc; exon 3/4-R:catttctctactagataaagagg.

In vitro translation

A reticulocyte lysate system and nonradioactive translation detection system (Promega) were used to translate AK045681 full-length and spliced cDNAs in vitro according to the instructions from the manufacture. Briefly, an equal concentration (1μg) of each cDNA was added to the reticulocyte lysate transcription/translation system with biotinylated lysine t-RNA molecules. The reactions were incubated at 30°C for 1 hour. After the reactions were completed, the products were ran in SDS-PAGE gel and transferred into the PVDF membrane. The membrane was incubated in TBST for 60 minutes followed by incubation with streptavidin-AP at room temperature for another hour. After the completion of streptavidin-AP binding, the membrane was treated with the substrate for alkaline phosphate until the bands of interest emerged.

X-gal staining

For the whole mount X-gal staining the brains were taken out at the appropriate time of development. The brains were incubated in the LacZ fixative (1% formaldehyde, 0.2% gluteraldehyde, 2mM $MgCl_2$, 5mM EGTA, 0.02% Nonidet P40 (NP40) in PBS) for 20 minutes at 4°C. After fixation, the brains were washed three times (20 minutes each) with PBS at room temperature. After the wash, the brains were incubated in the staining solution (5mM potassium ferricyanide, 5mM potassium ferrocyanide, 2mM $MgCl_2$, 0.01% sodium deoxycholate, 0.02% NP-40, 1mg/ml X-Gal in PBS) and kept in the dark at 4°C overnight. After the staining was completed, the brains were washed twice (15 minutes each) with PBS at room temperature and stored in 70% ethanol until analyzed.

For the X-gal slide staining, the brains were fixed for 4 hours with rocking at 4°C. After the fixation, the brains were washed three times with the wash buffer following by three washes with PBS. Then the brains were cryoprotected in 15% sucrose/PBS for 1 hour followed by the cryoprotection in 30% sucrose/PBS overnight at 4°C. The samples were embedded into the embedding medium (Tissue-Tek; Sakura Finetek) and sectioned at 10-μm thickness. The sectioned samples were incubated in the LacZ staining solution for 4-6 hours at 37C and washed three times in PBS. Finally the sectioned samples were dehydrated from 50% to 100% ethanol and cover slipped.

Immunostaining

Immunostaining was carried out on 10-μm sections using the following primary antibodies: anti-β-galactosidase (rabbit & chicken; 1:250; Abcam), anti-NeuN (mouse; 1:1000; Millpore), anti-GFAP (mouse; 1: 250; Covance), anti-MBP (mouse; 1:250; Covance), anti-calbindin (mouse; 1:250; Swat), anti-GABAaRα6 (rabibit; 1:500; Millipore), anti-GABA (g-pig; 1:500; Abcam). Primary antibodies were detected with secondary antibodies labeled with Cy3, or FITC (1:1000; Invitrogen). Primary antibodies used for immunostaining of the cultured neurons and mouse embryonic fibroblasts were the following: mouse monoclonal anti-HA (1:1000, Covance), mouse monoclonal anti-Tuj1 (1:200, Covance), rabbit polyclonal anti-SmB/B'/N (1:500, Santa Cruz biotechnology). Primary antibodies were detected with secondary antibodies labeled with Cy3 or FITC (1:1000; Invitrogen). DAPI was used for the nuclear staining.

Behavioral analyses

Metric spatial processing task

The protocol was adapted from Goodrich-Hunsaker, et al., (2005, 2008b). The metric spatial processing task was performed on a circular table measuring1m in diameter. The table was covered with a clean, clear plastic cover. Four objects were used as stimuli in the task (small bottle filled with red liquid, an overturned coffee mug, a large pencil stand, and a floating orange duck). Those objects measured between 2.5-5 cm at the base and between 5 and 15 cm tall, and were chosen to be textually and visually unique and easy to distinguish for mice. The table was surrounded on four sides by walls (at a distance of ~1 m) with discrete distal cues available at the same level as the table surface affixed to the wall. The mice were placed back in their home cage between the habituation and test period. The circular table was wiped down with 70% ethanol after testing of each mouse to remove any trace of olfactory cue that could affect object exploration by mice tested later. The two objects were placed 45 cm apart on the circular table. The mouse was allowed 15 minutes to freely explore the tabletop, stimulus objects on the table and distal environmental cues. After a 15-minute habituation period, the mouse was placed back in the home cage for 5 minutes. During the intermission the objects were moved closer to each other so that the distance between the objects was 30 cm. The mouse was placed back onto the table again to re-explore the objects for 5 minutes of the testing period.

Topological spatial processing task

The procedure of the topological spatial task was described elsewhere (Goodrich-Hunsaker, et al., 2005, 2008b; Lee et al., 2009). The topological spatial processing task was performed on the behavioral apparatus described above. Two novel objects that were different from those used in the metric spatial processing task were used as stimuli. Mice were placed on the table and allowed to explore exactly for 15 minutes as in the metric spatial processing task. During the intermission of the experiment the positions of objects on the table were switched (i.e., the object on the left in the habituation period was now placed on the right and vise versa). The spatial locations of two objects were the same as in the habituation period but the positions of two objects were exchanged in the testing period. For both the metric and topological spatial processing task, time that mice spent to explore each object was recorded. The operational definition of exploration was that mice actively sniffed or touched the objects with the nose, vibrissa or forepaws.

Rota-rod motor learning task

The Rota-rod motor learning task is described elsewhere (Homles et al., 2001). The Rota-rod motor learning task consisted of four trials (three training sessions plus pre-training) per day over 4 consecutive days. Each training trial lasted for 5 minutes, and the intertrial interval was for 20 minutes. Mice were acclimated to the behavioral room for 30 minutes before the task began. As the pretraining session, the rotation speed of the rod was set to 4rpm constant speed. The mice were placed onto the rod and must keep their balance for 60 seconds. The mice were put back onto the rod when they fell off the rod before the time expired. The mice must stay on the rod for 60 seconds successfully to

pass the pretraining. The purpose of the pretraining was to get them accustomed to the task. After the pre-training session, the mice were placed back to their home cages and rested for 30 minutes. After a 30-minute break, the training sessions began. The mice were put back onto the rod, which was set to accelerate from 4rpm to 35rpm in 300 seconds. The mice were moved back into their home cages after they fell off the rod. The latency for the mice to fell off the rod was recorded. The Rota-rod apparatus was wiped clean with 70% ethanol and dried well between the trainings.

Elevated plus maze

The protocol of the elevated plus maze was described elsewhere (Walf & Frye et al., 2007; Holmes et al., 2002). Briefly, the elevated plus maze consists of four arms (two open arms without walls and two closed arms with 15.25 cm high walls). Each arm of the maze is 30 cm long and 5 cm wide and is attached to solid metal legs, which elevate the maze 60 cm off of the stage that the maze is on. Thirty minutes before the behavioral tests were performed, mice were transferred to the behavioral test room and acclimated to the room where the test was administrated. Mice were placed in the interaction of the four arms of the maze facing an open arm. The exploratory behavior of mice was detected by a video-tracking system (Any-Maze, Stoelting) for 5 minutes. Entries and durations in each arm (open, closed arms and intersection) were recorded by the video tracking system. The video tracking system also recorded distance that mice travelled in each arm. The maze was wiped clean with 70% ethanol and dried well before testing with another mouse.

Results

AK045681 is knocked out by the *piggybac* DNA transposon system

RT-PCR analysis and northern blotting were performed to identify the presence of AK045681 transcripts in the AK045681 KO mice. RT-PCR showed that AK045681 mRNA was not present in the brains of the AK045681 null mutant mice when mRNA of AK045681 was amplified by using the primer pair designed for the exon3-exon4 sequence (Figure 3-1). On the contrary, RT-PCR analysis demonstrated the presence of AK045681 mRNA in the WT littermate mice when AK045681 mRNA was amplified using the same primers above (Figure 3-1). Because the open reading frame (ORF) of AK045681 is located only in exon 4, RT-PCR analysis was performed to see if the AK045681 transcripts with exon4 were eliminated in the AK045681 null mutant mice. The results of RT-PCR unexpectedly showed the presence of mRNA of exon4 in the AK045681 mutant and the WT littermate mice, although the amplified band for the mutant mice was much less intense than the one for the WT mice (Figure 3-1). To verify the results of the RT-PCR, the northern blotting was performed by using probes for exon 3-4 and exon 4 only. The results showed the presence of both exon3/4 or exon4 of AK045681 in the WT littermate mice, but neither transcript was present in the AK045681 KO mice (Figure 3-1).

Our *piggybac* DNA transposon cassette carries IRES-LacZ, which is expressed as AK045681 is. Therefore, LacZ serves as the marker of AK045681. Whole mount X-gal staining of the brains of AK045681 mutant mice revealed the timing of AK045681 expression in brain. As Figure 3-2 shows, AK045681 started to express on postnatal day 12 and its expression persisted into adulthood.

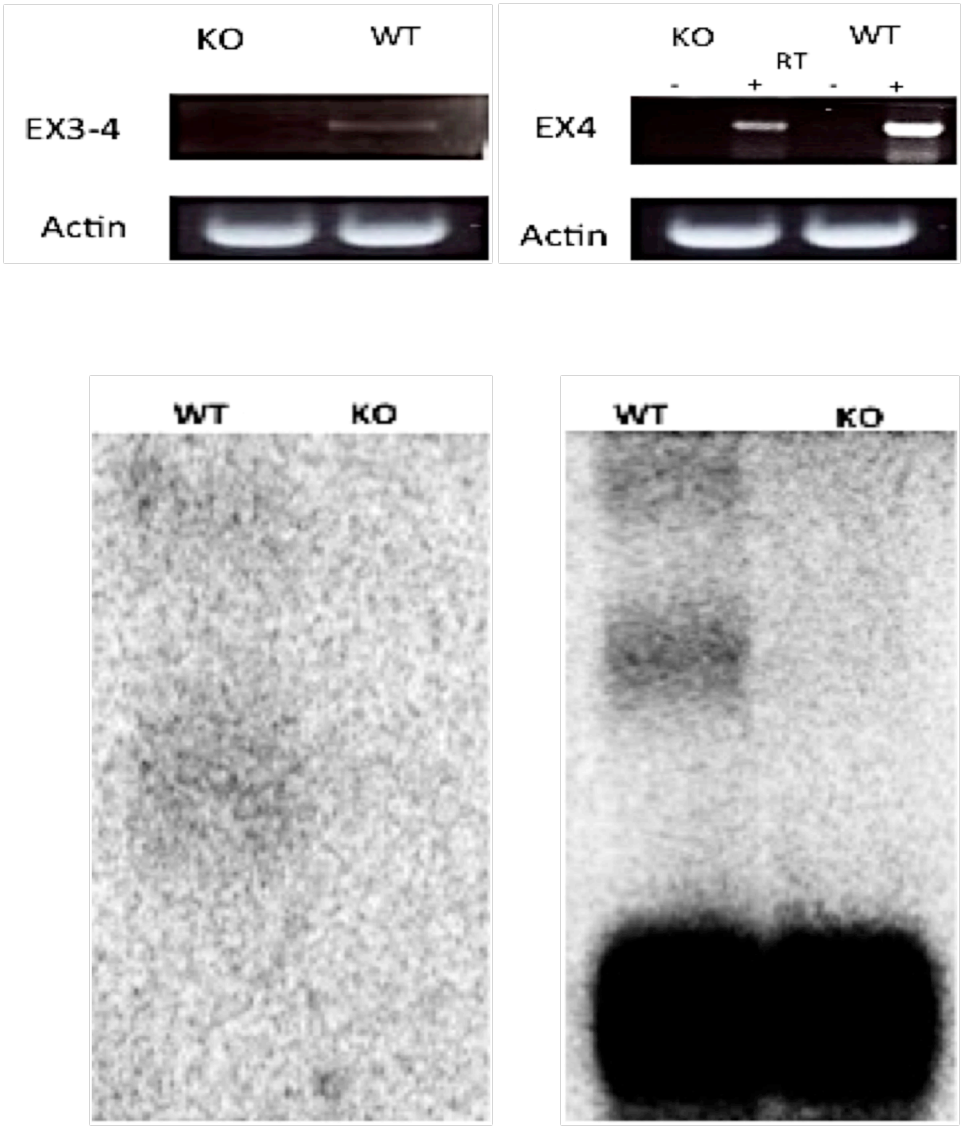

Figure 3- 1. RT-PCR and northern blotting of AK045681 transcripts in mouse brain. RT-PCR analyses (Top-left panel) showed that no AK045681 transcript containing Ex3/4 was detected whereas a little trace of AK transcripts with ex4 (ORF-harboring) in AK045681 null mutant mice. Northern blotting (bottom panels) showed no detection of both AK transcripts (ex3/4 and ex4 only). AK045681 is expressed postnatally (postnatal day 12) in the hippocampus, amygdala, cortex, cerebellum, hypothalamus and anterior olfactory nucleus.

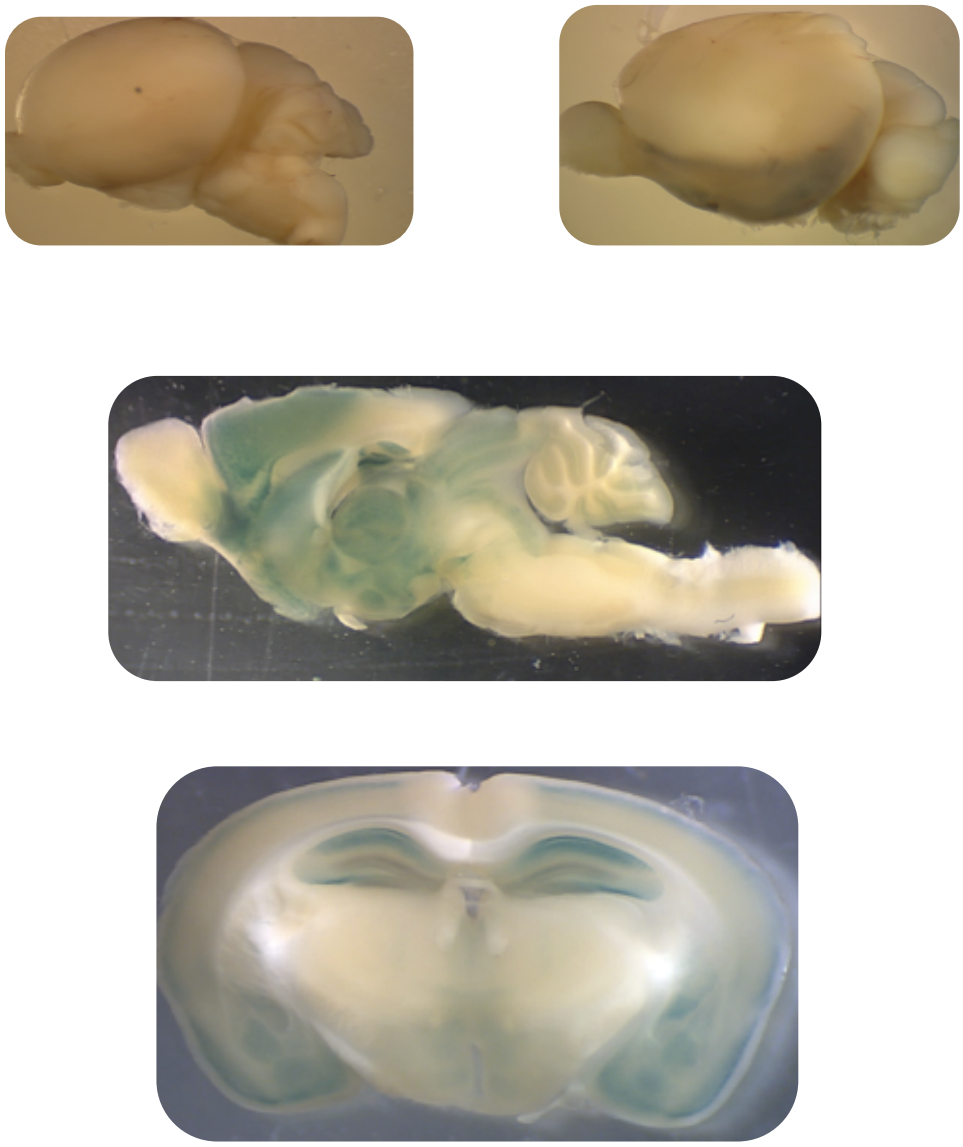

Figure 3- 2. Whole mount X-gal staining of the brains of the AK045681 null mutant mice. X-gal staining was detected in the whole brain of the AK045681 null mutant mice at postnatal day 12 (Top-right) whereas no staining was detected in the whole brain of the AK045681 null mutant mice. The sagittal (middle-panel) and coronal (bottom-panel) view of the whole brain showed the X-gal staining in specific neural regions in the brains of the AK045681 null mice.

The X-gal staining of the whole brains showed that AK045681 was expressed in specific brain regions such as the cortex, hippocampus, amygdala, olfactory system, and the cerebellum. The X-gal staining of individual brain sections showed more cellular detailed expression of AK045681. The strong expression of AK045681 was observed in the CA regions of hippocampus. The superficial layers of the cortex showed the intense X-gal staining. In the cerebellum, the inner granular layer and the Purkinje cell layer showed the expression of AK045681. The X-gal staining confirmed the strong expression of AK045681 in the brain regions including the amygdala. Additionally, the expression of AK045681 was observed in the striatum, which was not observed on the whole mount X-gal staining of the brains (data not shown).

AK045681 is expressed and localized in the nucleus of GABAergic neurons in the mouse brain

As Figure 3-3 shows, co-localization of AK045681 with NeuN was detected in the brains of the AK045681 null mutant mice. The expression of AK045681 was mainly observed in pyramidal cell layer of the hippocampus. Co-localization of AK045681 and GABA was detected in the pyramidal cell layer of the hippocampus. Double immunostaining of AK045681 and GABA was also found in other AK-expressing brain regions including the cortex, the anterior olfactory nucleus, the striatum (data not shown), and the amygdala (data not shown). Co-localization of AK045681 and Purkinje cell specific marker Calbindin validated the expression of the AK045681 in the Purkinje cells of the cerebellum. Furthermore, the cerebellar granular cell specific marker, GABAaR

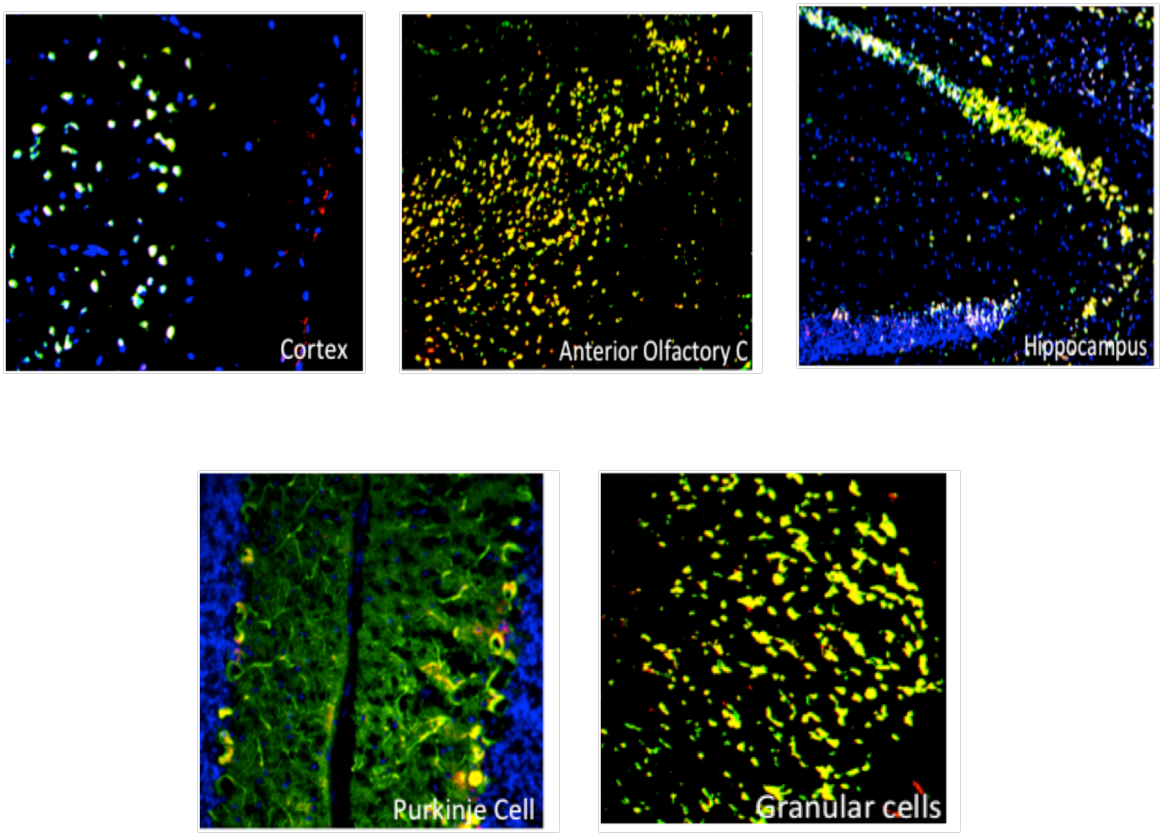

Figure 3- 3. AK045681 is expressed in the GABAergic neurons in different neuronal substrates in the mouse brain. Co-localizations of β-gal (Green) and GABA (Red) were detected in the cortex (Top-left), the anterior olfactory nucleus (Top-middle), and the hippocampus (Top- right). AK045681 is also expressed in Purkinje cells (Green: Calbindin, Red: β-gal, Bule; DAPI) and granular cells (Green: GABAaR α6, Red:I β-gal) in the cerebellum.

subunit α6, was co-localized with AK045681, which confirmed the expression of AK045681 in the granule cells in the inner granular cell layer of the cerebellum.

The amino acid sequence of AK045681 is similar to the spliceosomal protein U1C. This suggests that AK045681 may play a role in RNA splicing and maturation. RNA splicing by the spliceosome occurs in the nuclei of cells. Localization of AK045681 in the nucleus of cell would be the further indication that the AK045681 may be involved in RNA splicing and maturation. Co-localization of HA-tagged AK045681 protein and DAPI nuclear staining demonstrated the localization of AK045681 in the nucleus of the mouse fibroblasts transfected with HA-tagged-AK045681 ORF plasmids (Figure 3-4). Furthermore, the co-localization of AK045681 with DAPI nuclear staining was also observed in the nucleus of the HA-AK045681-ORF transfected mouse cortical neurons (Figure 3-4). Additionally, the nuclear spliceosomal protein, SmB/B'/N, was used to validate the nuclear localization of AK045681. Triple immunostainings of HA-tagged AK045681, SmB/B'/N, and DAPI, showed nuclear co-localization of AK045681 in mouse fibroblasts (Figure 3-4).

AK045681 is translatable in vitro

AK045681 has two forms of its transcripts, a full-length and a splicing variant, which is the lack of exon 2-3 of AK045681. To determine whether the transcripts of AK045681 are indeed translatable in vitro, cDNAs of those two forms of AK045681 were transcribed and translated in the reticulocyte lysate. As Figure 3-5 shows, both the full-length and the spliced variant of AK045681 were translated in vitro, which indicated that the proteins of AK045681 would possibly have the biological function in vivo.

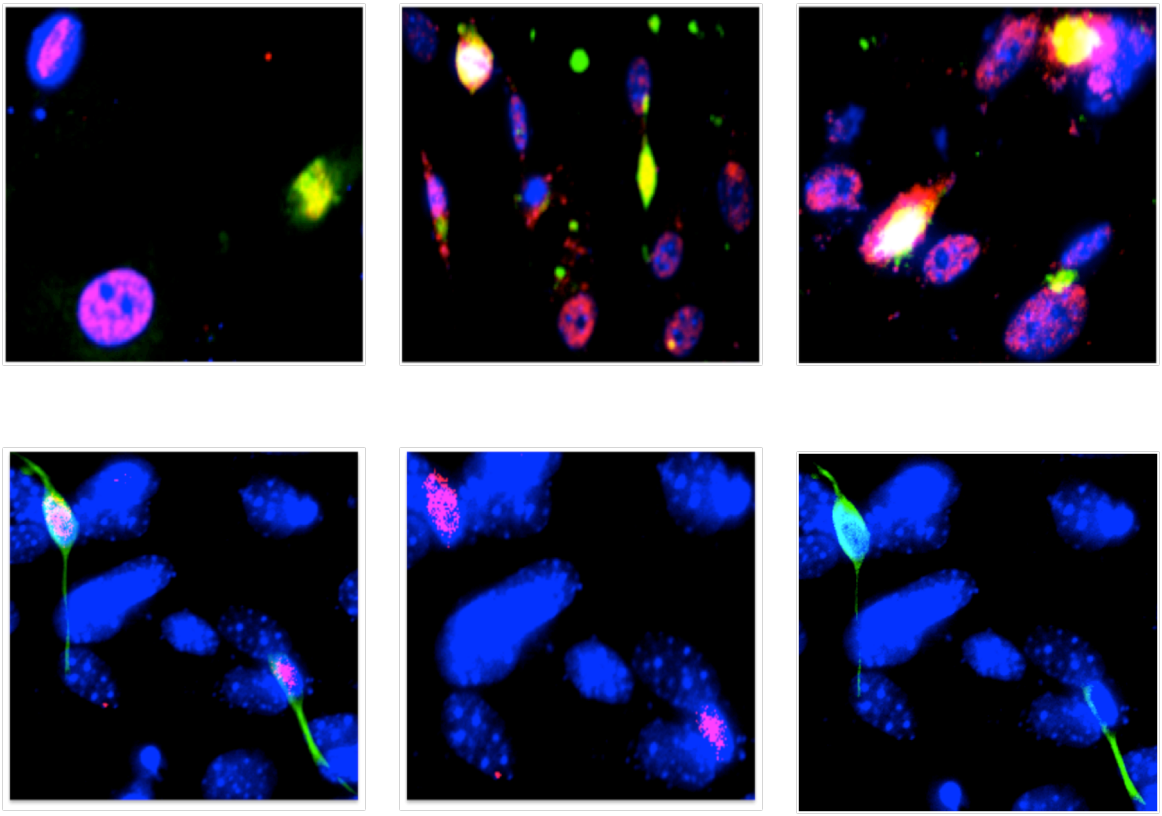

Figure 3- 4. The nuclear localization of AK045681 in the mouse fibroblasts (MEFs) and cortical neurons. In MEFs (Top panels) the HA-tagged AK045681 (Green: HA) co-localized with the nuclear spliceosomal protein SmB/B' (Red) and DAPI (Blue). Co-localization of the HA-tagged AK045681 and the nuclear staining DAPI were observed in the Tuj1-positive neurons (bottom panels; Green: Tuj1, Red: HA, Blue: DAPI)

Additionally, the results revealed that there was a noticeable difference in the level of the translated proteins between the full-length and the spliced AK045681 (Figure 3-5).

AK045681 was not a part of the spliceosome and was not involved in RNA splicing of GABA receptor and transporter genes

Biological functions of the translatable AK045681 protein were examined. The fact that AK045681was translatable in vitro and its protein was localized in the nucleus of the neurons led me to look at the potential function of AK045681 as a spliceosomal protein. As stated earlier, the amino acid sequence of AK045681 is similar to the sequence of the spliceosomal U1C, which was shown to physically interact with another spliceosomal protein U1-70K. Potential function(s) of the translated AK045681 protein as an alterative spliceosomal protein to U1C in the spliceosomal complex were examined. Co-immunoprecipitation of HA-tagged AK045681 to U1-70K in HEK293T cells revealed no physical association between the two proteins, which indicated AK045681 was not the part of the spliceosome (Figure 3-5).

The expression of AK045681 in the GABAergic neurons led me to look at the mRNA splicing activity of the GABA receptor and transporter genes. The results of RT-PCR showed normal RNA splicing profiles for GABA receptor and transporter genes (Figure 3-5). Therefore, AK045681 was not involved in mRNA splicing of the GABA receptor and transport genes.

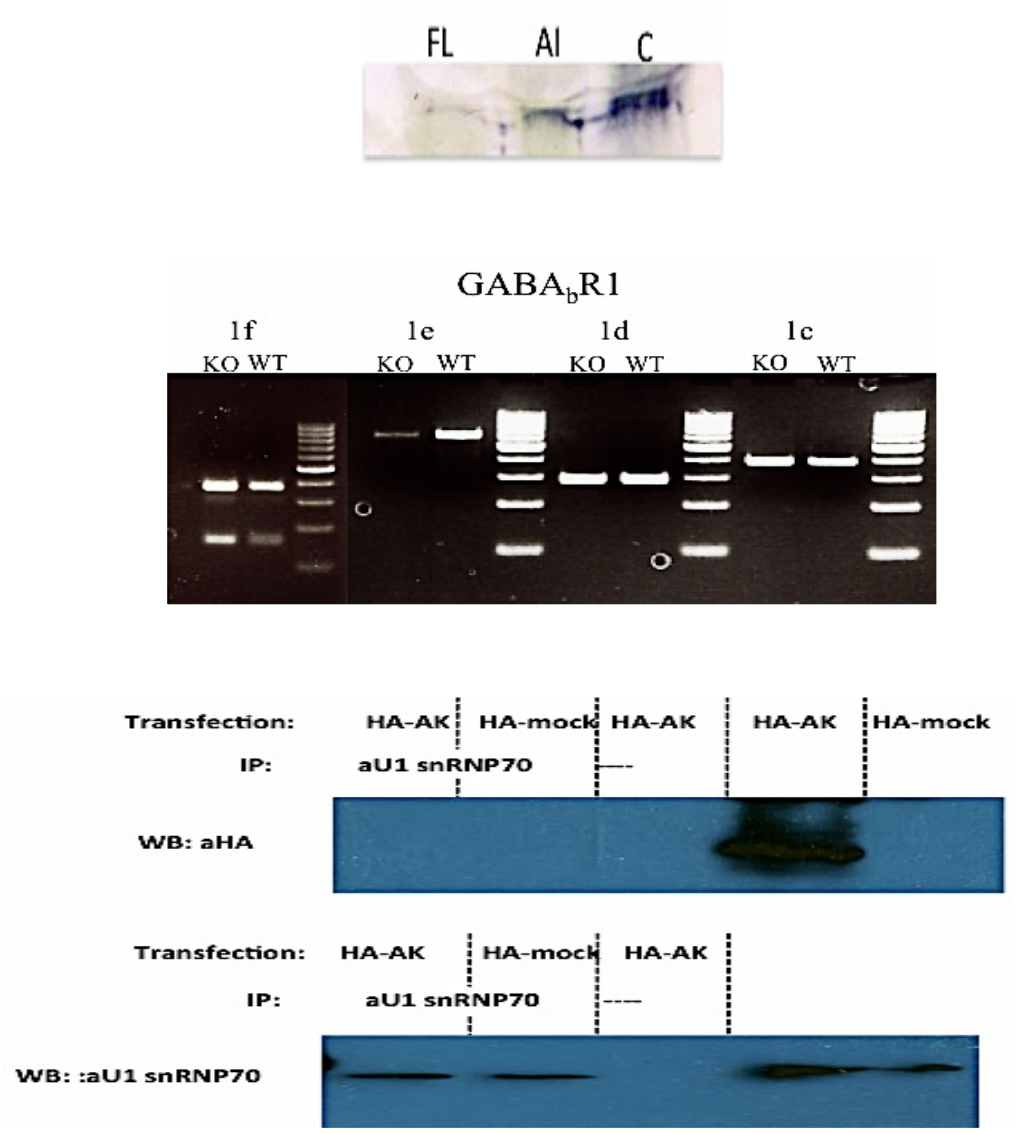

Figure 3- 5. Biochecial analyses of AK045681. In vitro translation of a full-length cDNA (FL) and alterantive isform (AI) of AK045681 (Top panel). The ORF of AK045681 was translated as control (C). (Middle panel): mRNA profiles of GABA receptors genes (Middle panel). Co-immunoprecipitation of HA-tagged Ak045681 protein with the nuclear spliceosomal protein U1 snRNP70 (Bottom panel).

Spatial and motor learning is intact in AK045681 mutant mice

AK045681 KO mice are viable. Nissl staining showed that gross brain morphology of AK045681 appeared to be normal (data not shown). The timing (postnatal day 12) and major locations (the hippocampus, the cerebellum and the amygdala) of the gene expression in the brain suggested that AK045681 might be involved in regulating learning, memory and other behaviors. The metric and the topological spatial learning task were used to examine the function of the hippocampus. As Figure 3-6 shows, there is no significant difference in time that both AK-KO mice and their WT littermate explored the novel objects in the metric and the topological spatial processing tasks.

Rota-rod motor learning task was used to assess the function of the cerebellum in AK045681 mutant mice. There was no significant difference in the average latency to fall off the rod between the AK045681 mutant and the WT littermate mice in each trail of every training days (Figure 3-7). The average latency to fall off the rod in each day between the AK045681 mutant and the WT littermate mice was compared, and there was no significant difference between the groups (Figure 3-7). These results indicate that the functions of the hippocampus and the cerebellum are intact in the AK 045681 mutant mice.

Disruption of AK0456821 interferes anxiety-related behaviors

AK045681 is strongly expressed in the limbic system such as the hippocampus and the amygdala. It has been shown that the limbic system plays crucial roles in regulating fear-related behaviors including the anxiety behavior. We subjected the AK045681 KO mice and their WT littermate mice to the elevated plus maze task to

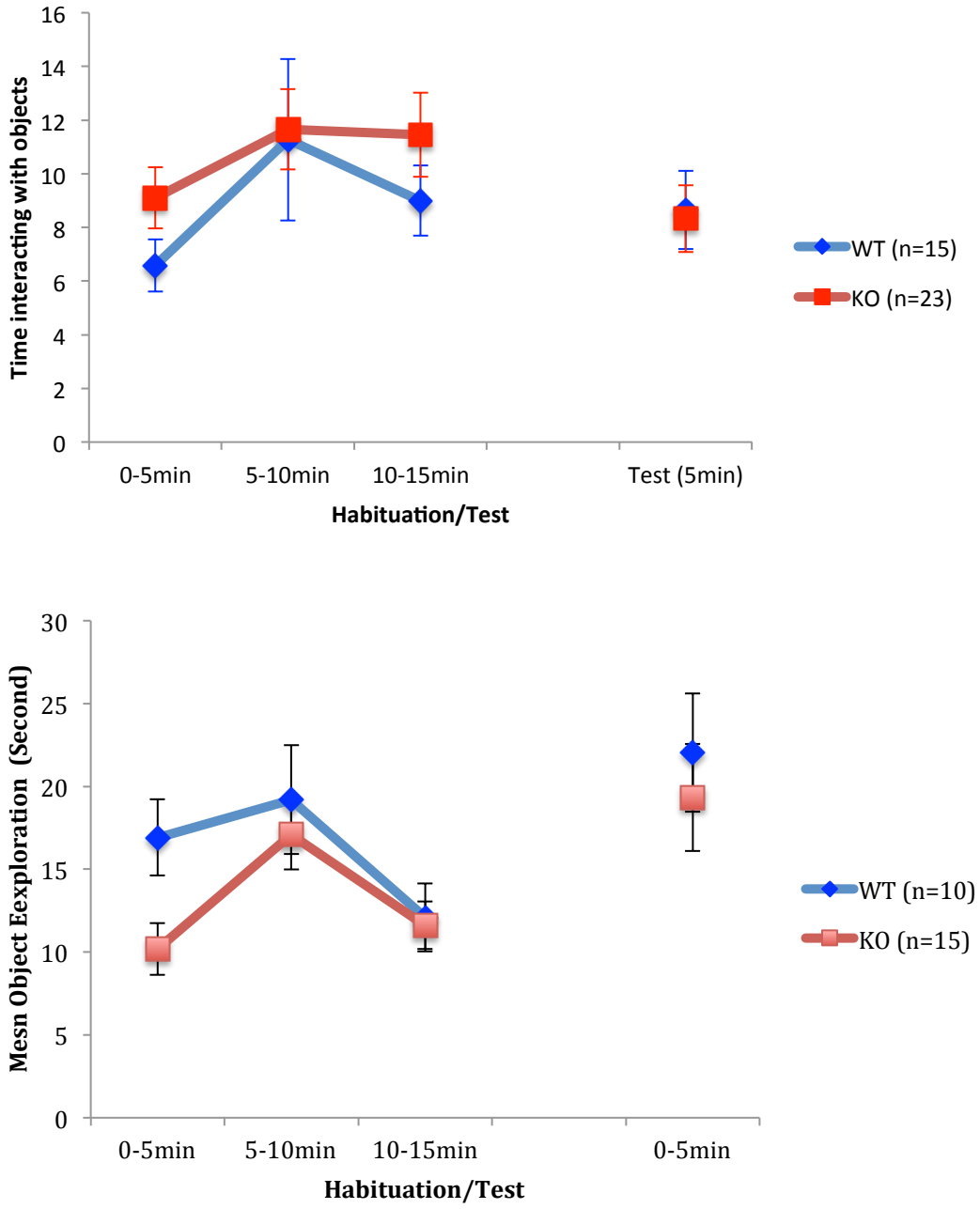

Figure 3- 6. Spatial (hippocampal-dependent) are intact in AK045681 null mutant mice. Mean time that AK-KO mice and their WT littermates interacted with the novel objects on the testing arena during a 15 minute habituation and a 5 minute testing session in metric spatial task (Top). Mean time that AK-KO mice and their WT littermates interacted with the novel objects on the testing arena during a 15 minute habituation and a 5 minute testing session in the topological spatial task (Bottom). N=12 and 23 for WT and AK-KO, respectively for these tasks.

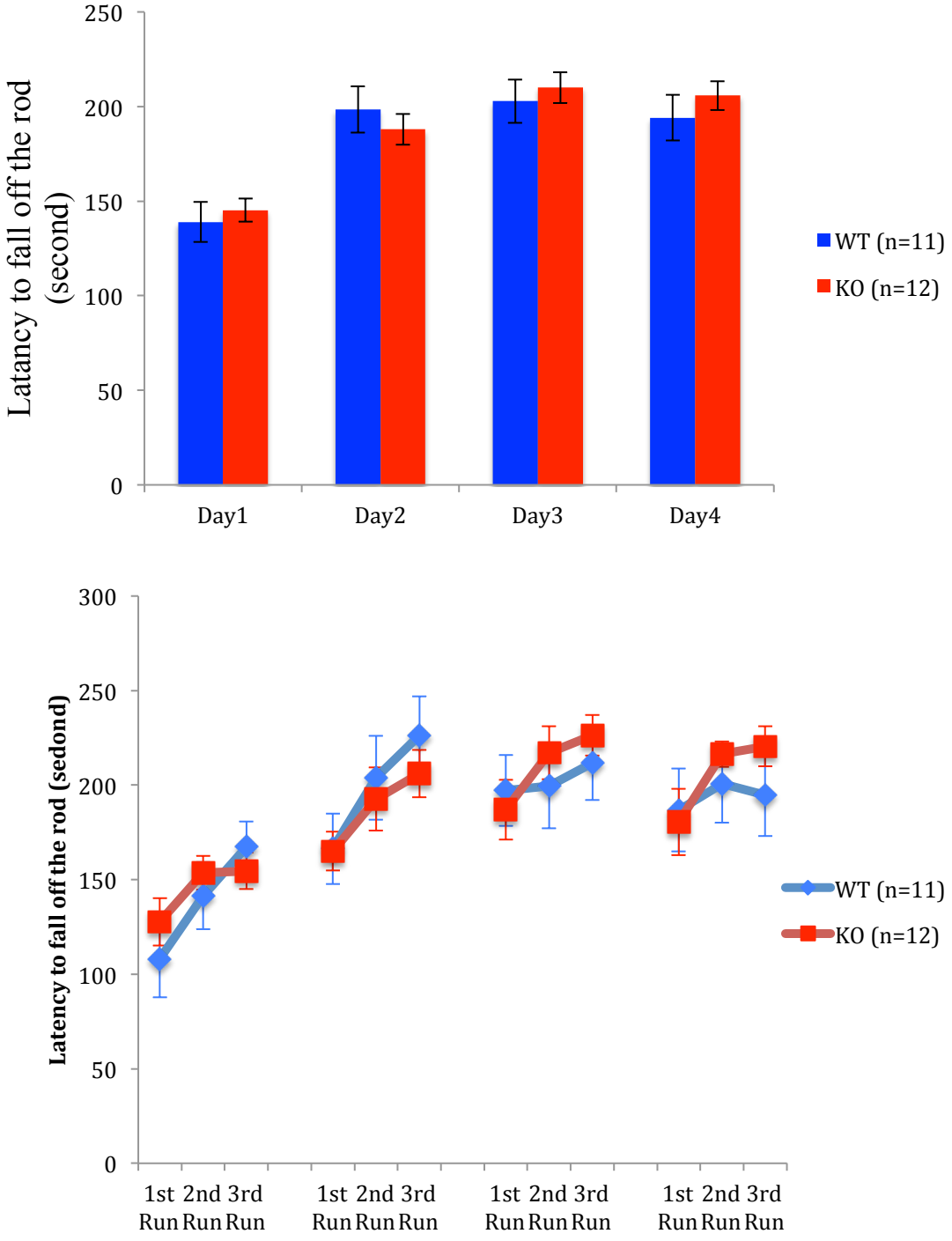

Figure 3-7. Mean time latency to fall off the rod between the AK-KO mice and their WT littermate mice during a single 5 minute-trail in each day in the Rota-rod task (Top). Mean time latency to fall off the rod between the AK-KO mice and their WT littermate mice during each single 5 minute-trail in the Rota-rod task (Bottom). n=12 and 11 for the AK-KO mice and the WT mice, respectively.

examine the functions of the limbic system. There was no difference in the total number of arm entries during a single 5-minute session between the AK045681 KO and their WT littermate mice. When the time spent in each arm was compared, the AK045681 KO mice spent significantly more time in the open arms than their WT littermate mice did. On the contrary, the AK045681 KO mice spent significantly less time in the closed arms than their WT littermate mice did (Figure 3-8). There was no significant difference in the time spent in the intersection between the AK045681 mutant and their WT littermate mice (Figure 3-8). Finally, there was no significant difference in the distance travelled in the closed arms and the intersection between the AK045681 KO and their WT littermate mice. However, when the distance travelled in the open arms between the AK045681 mutant mice and their littermate mice was compared, the AK045681 KO mice traveled significantly more distance in the open arm than their WT littermate mice did (Figure 3-8). The results showed that the AK045681 mutant mice showed significantly less anxiety than their WT littermate mice, and disruption of the AK045681 impairs the regulation of anxiety-related behavior in mice.

Discussion

Characterization of the novel gene AK045681 was described in this study. The results presented here demonstrated that the novel gene AK045681 was postnatally expressed in the GABAergic neurons in the specific regions of the brain such as the hippocampus, the amygdala, the cortex, and the hypothalamus. The AK04568 null mutant mice showed reduced levels of the anxiety-related behavior in the elevated plus maze. The results suggest that AK045681 may play a role in regulating an anxiety-

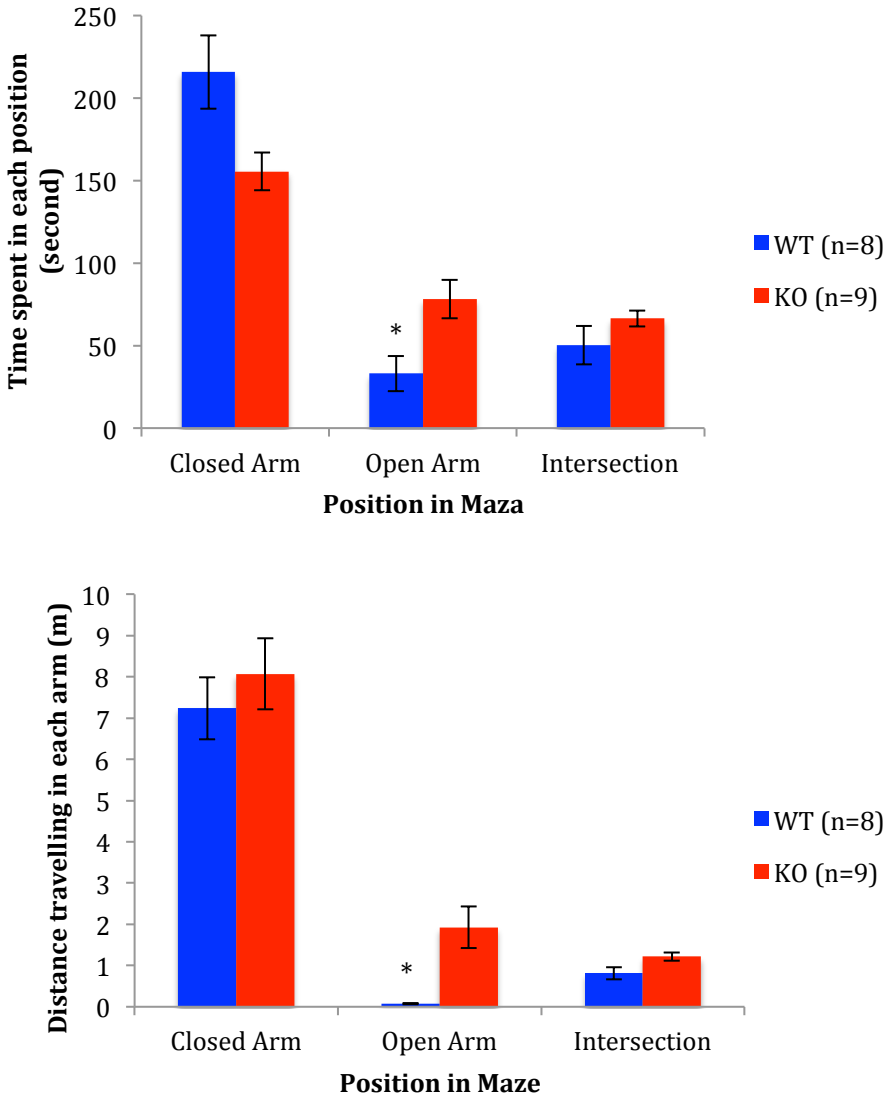

Figure 3- 8. AK null mutant mice showed significant reduction of expression of anxiety-related behavior. Mean time that the AK-KO mice and their WT littermate mice spent in each arms (open, closed, or intersection) during a 5 minute-test in the elevated plus maze task (Top panel). Mean distance that AK-KO mice and their WT littermate mice traveled in each arms during a single 5 minute-test in the elevated plus maze (Middle panel). Mean number of entries that AK-KO mice and their WT littermate mice made into each arm during a single 5 minute-test in the elevated plus maze (Bottom panel). N=9 and 7 for the AK-KO mice and the WT mice, respectively. The asterisk (*) indicates significant difference when $p < 0.05$.

Figure 3-8 Cont.

related behavior through the GABA-mediated neurotransmission. Imaging studies in humans, as well as lesion and microinfusion studies in animals, have indicated that brain regions such as the hippocampus, the amygdala, the cortex, and the hypothalamus were crucial neuronal substrates for regulation of fear and anxiety behaviors (Charney & Drevets, 2002). Pharmacological evidence has suggested that endogenous GABA plays a key role in regulation of anxiety since the drugs that altered GABAergic neurotransmission, such as benzodiazepepines, have been successfully managing the symptoms of anxiety disorders in humans (Goman et al., 2002; Mohler, 2007; Clement et al., 2002). Furthermore, evidence from the knockout and transgenic mice that exhibited abnormal levels of anxiety-related behaviors supported the potential role of GABA-mediated neurotransmission in the pathophysiology of anxiety. For instance, the mice with the gamma2 subunit of GABAa receptor mutation showed increased anxiety behavior in both the elevated plus maze and the light and dark transition task. Alternative splicing creates two forms of transcripts of the gamma2 subunit of GABAa receptor, γ2L and γ2s. Interestingly, the mice with the mutation of γ2L of GABAa receptor showed an increase in anxiety behavior in the elevated plus maze task (Mohler, 2007; Clement et al., 2002). The mice deficient in the GABA-synthetic gene, GAD-65, showed a reduction in GABA neurotransmission and exhibited increased levels of anxiety-related behavior in the elevated plus maze and the light and dark transition task (Kash et al., 1999). The GAD-65 knockout mice were also less sensitive to the acute effects of benzodiazepepines and barbiturates (another positive modulator of GABAergic neurotransmission) (Kash et al., 1999). In addition, GAD-65 has been associated with the panic disorder-related trait of behavioral inhibition in children (Smoller et al., 2001).

Amino acid sequence similarity of this novel gene to the spliceosomal protein U1C suggests that functional roles of this novel gene may be RNA splicing. AK045681 may be a part of the major spliceosome or a splicing factor to regulate RNA splicing and maturation. Furthermore, the potential function of AK045681 in RNA splicing and maturation may be activity-dependent. It has been reported that some neuronal genes such as BDNF produce multiple forms of its proteins in the activity-dependent manners (e.g. when some specific learning occurs) by utilizing the alternative splicing mechanism (reviewed in Li et al., 2007). I hypothesized that AK045681 might be involved in RNA splicing and maturation (either activity-dependent or not) of the genes in GABA-mediated neurotransmission. Therefore, the disruption of the splicing mechanism in the brain regions such as the hippocampus, the amygdala, the cortex and the hypothalamus could interfere with the regulation of anxiety-related behavior.

To test this hypothesis, three molecular requirements to be a part of the spliceosome were examined. First, since mRNA splicing takes place in the nucleus of the cells, the protein of this novel gene should be localized in the nucleus of the neurons if the function of AK045681 was RNA splicing as suggested. The results showed the co-localization of the immune-tagged (HA) AK045681 protein and the nuclear DAPI staining in the neurons. This novel protein was indeed localized in the nucleus of the neurons. Nuclear co-localization of AK045681 with the spliceosomal proteins further indicates that AK045681 might be involved in RNA splicing in the nucleus of the neurons. Indeed, AK045681 co-localized the spliceosomal protein, SmB/B'/N with the HA-tagged AK045681 protein in the nucleus of the neurons, as the results presented have showed.

Ultimately, if AK045681 is an activity-dependent alternative RNA splicer, AK045681 might show a physical interaction (or binding) with the spliceosomal proteins. Because the amino acid sequence of this novel gene is similar to the spliceosomal protein U1C, AK045681 might form the spliceosome complex and function as U1C does. In the spliceosome complex, U1C physically interacts with the spliceosomal proteins U1-70K and SmB/B'/N to recognize a 5' splice site (Bedford et al., 1997). Furthermore, if AK045681 is truly an alternative splicer, the target gene(s) that AK045681 alternatively splice may be the gene(s) playing roles in GABA-mediated neurotransmission. Thus, potential roles of AK045681 in a part of the spliceosome as an alternative part for U1C, I examined the physical interaction of AK045681 with the spliceosomal proteins U1-70K or SmB/B'/N by co-immunoprecipitation. In addition, alternative splicing mRNA profiles of the GABA receptors and transporters genes were also analyzed to test this hypothesis. As the results presented here, however, AK045681 did not interact with either spliceosomal proteins, and it appeared to be normal mRNA splicing profiles of these genes between the AK045681 mutant mice and their wild-type littermates.

Alternative splicing created two transcripts of AK045681. The full-length of AK045681 consisted of exon1 through 4, whereas the isoform was made of only exon1 and exon4. The difference between those transcripts of AK045681 lay in 5' untranslated region (5'UTR) of the gene; short 5'UTR (~900-bp) for the spliced and long 5'UTR (~1100-bp) for the full-length transcript. As the results displayed, the spliced transcripts with the short 5'UTR translated efficiently, while the full-length transcripts with the long 5'UTR poorly translated when both AK045681 transcripts were translated in vitro. It has been reported that the genes yield mRNAs containing distinct 5'UTRs, and those distinct

5'UTRs function as translation regulators in vivo. RUNX1 is a member of the runt domain transcription factors that plays critical regulatory roles in hematopoiesis and osteogenesis (Kumano & Kurokawa, 2010:Ito & Miyazono, 2003:Speck, 2001). RUNX1is expressed through two promoters known as distal (D) and proximal (P), which produce two different 5'UTRs (D-UTR and P-UTR) (Ghozi et al., 1996). When D-UTR and P-UTR were translated in vivo, P-UTR and D-UTR were equally translated. The longer P-UTR was found to be weakly translated, while the shorter D-UTR was strongly translated (Pozner et al., 2000). The difference in the translation efficiency is due to the length and *cis*-elements that lie in each UTR. Translation through D-UTR was found to be cap-dependent in contrast P-UTR was facilitated through cap-independent and internal ribosomal entry site (IRES)-dependent translation (Pozner et al., 2000). It has been suspected that IRES-mediated translation could be a part of regular translation mechanism; however, mRNA with IRES could be translated only under certain cellular circumstances, such as growth arrest, serum starvation, γ-irradiation, hypoxic condition, where cap-dependent translation is stalled (Holick et al., 1999; Rhoads, 1993; Song et al., 1999). One could speculate that AK045681 would utilize a similar mechanism to ensure its gene expression under any biological circumstance.

The timing and pattern of AK045681 expression may be useful in terms of genetic engineering. As stated earlier, this novel gene is turned only on postnatally and its expression persists throughout adulthood. Furthermore, AK045681 is only expressed in the GABAergic neurons in the specific neuronal regions including the hippocampus and the amygdala. There are a few Cre drivers that truly perform the recombination only in the postnatal day. AK045681 may be a terrific gene to be utilized as a postnatal

neuronal-specific Cre driver. Because AK045681 is only expressed in GABAergic neurons in the brains, this will add more specificity for its usage as a postnatal neuronal-specific Cre driver.

The biological functions of AK045681 still remain unknown. To uncover the biological role of AK045681 will provide valuable information to the genetic and molecular mechanisms of anxiety regulation.

References

Abovich, N. and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell, 89, 403-412.

Bedford, M.T., Chan, D.C. and Leder, P. (1997). FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. EMBO J., 16, 2376-2383.

Cadinanos, J. and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. Nucleic Acids Res., 35, e87.

Carlson, C.M. and Largaespada, D.A. (2005). Insertional mutagenesis in mice: new perspective and tools. Nat Rev Genet., 6, 568-580.

Clement, Y., Calatayud, F. and Belzung, C. (2002). Genetic basis of anxiety-like behavior: a critical review. Brain Res Bull., 57, 57-71.

Copeland, N.G. and Jenkins, N.A. (2010). Harnessing transposons for cancer gene discovery. Nat Rev Cancer., 10, 696-706.

Diaz, E. (2009) From microarray to mechanisms of brain development and function. Biochem Biophys Res Commun., 385, 129-131.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y. and Xu, T. (2005). Efficient transposition of the *piggyBac* (PB) transposon in mammalian cells and mice. Cell., 122, 473-483.

Elick, T.A., Bauser, C.A. and Fraser, M.J. (1996). Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. Genetica. 98, 33-41.

Forch, P., Puig, O., Martinez, C., Seraphin, B. and Valcarcel, J. (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. EMBO J., 21, 6882-6892.

Fraser, M.J., Ciszczon, T., Elick, T. and Bauser, C. (1996). Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cells lines from two species of Lepidoptera. Insect Mol Biol., 5, 141-151.

Ghozi, M.C., Bernstein, Y., Negreanu, V., Leabanon, D. and Groner, Y. (1996). Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. Proc Natl Acad Sci USA., 93, 1935-1940.

Goman, J.M., Kent, J.M. and Coplan, J.D. (2002). Current and emerging therapeutics of anxiety and stress disorders. Neuropharmacology: The fifth generation of progress. Philadelphia Lippincott Williams and Wilkins., pp. 967-980

Goodrich-Hunsaker, N. J., Hunsaker, M., & Kesner, R. P. (2005). Dissociating the role of the parietal cortex and dorsal hippocampus for spatial information processing. Behavioral Neuroscience*, 119*, 1307-1315.

Goodrich-Hunsaker, N. J., Hunsaker, M., & Kesner, R. P. (2008b). The interactions and dissociations of the dorsal hippocampus subregions: How the dentate gyrus, CA3, and CA1 process spatial information. Behavioral Neuroscience*, 122*, 16-26.

Holcik, M., Lefebvre, C., Yeh, T., Chow, T. and Korneluk, G. (1999). A new internal-ribosome-entry-site motif potentiates XIAP-mediated cytoprotection. Nat Cell Biol., 1, 190-192.

Holmes, A., Holon, T.R., Gleason, T.C., Llu, Z., Drelling, J., Sibley, D.R. and Crawley, J.N. (2001). Behavioral characterization of dopamine D5 receptor null mutant mice. Behav Neurosci., 115, 1129-1144.

Holmes, A., Yang R.J. and Crawley, J.N. (2002). Evaluation of an axiety-related phenotypes in galanin overexpressing transgenic mice. J Mol Neurosci., 18, 151-65.

Huang, X., Beullens, M., Zhang, J., Zhou, Y, Nicolaesuc, E., Lesage, B., Hu, Q., Wu, J., Bollen, M. and Shi, Y. (2009). Structure and function of the two tendem WW domains of the pre-mRNA splicing factor FBP21 (formin-binding protein 21). J Bio Chem., 284, 25375-25387.

Ingham, R.J. Colwill, K., Howard, C., Dettwiler, S., Lim, C.S.H., Yu, J., Raaijmakers, J., Gish, G., Mbamalu, G., Taylor, L., Yeung, B., Vassilovski, G., Amin, M., Chen, F., Matskova, L., Winberg, G., Ernberg, I., Linding, R., O'Donnell, P., Starostine, A., Keller, W., Metalnikov, P., Stark, C. and Pawson, T. (2005). Domains provide a platform for the assembly of multiprotein networks. Mol Cell Biol., 25, 7092-7106.

Ito, Y. and Miyazono K. (2003). RUNX transcription factors as key targets TGF-beta superfamily signaling. Curr Opin Genet Dev., 13, 48-54

Izsvak, Z and Ivics, Z. (2003). *Sleeping Beauty* transposition: biology and applications for molecular therapy. Mol Ther., 9, 147-156.

Ivics, Z., Hacket, P.B., Plasterk, R.H. and Izsvak, Z. (1997). Molecular reconstruction of *Sleeping Beauty*, a Tc-1-like transposon from fish, and its transposition in human cells. Cell, 91, 501-510.

Kash, S.F., Tecott, L.H., Hodge, C. and Baekkeskov, S. (1999). Increased anxiety and altered responses to anxiolytics in mice deficient in the 65-KDa isoform of glutamic acid decarboxylase. Proc Natl Acad Sci USA., 96, 1698-1703.

Killpack, K., Laufer, A., Mazzotta, J., Smith, R.D., Stevens, L.M., Stuber, C., Tan, L.R., Ventura, R., Woo, A., Zakrajsek, I., Zhao, L., Chen, F., Swimmer, C., Kopczynski, C., Duyk, G., Winberg, M.L. and Margolis, J. (2004). A complementary transposon tool kit for Drosophila melanogaster using *P* and *piggyBac*. Nat Genet., 36, 283-287.

Kofler, M., Heuer, K, Zech, T. and Freund, C. (2004). Recgnition sequences for the GYF domain reveal a possible spliceosomal function of CD2BP2. J Biol Chem., 279, 28292-28297.

Kolfer, M and Freund, C. (2006). The GYF domain. FEBS J., 273, 245-256.
Mates, L., Izsvak, Z. and Ivics, Z. (2007). Technology transefer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives. Genome Biol., 8, S1.

Kofler, M., Schuemann, M., Merz, C., Kosslick, D., Schlundt, A., Tannert, A., Schaefer, M., Luhrmann, R., Krause, E. and Freund, C. (2009). Proline-rich sequence recognition. Mol Cell Proteomics., 8, 2461-2473.

Komano, K. and Kurokawa, M. (2010). The role of RUNX1/AML1 and Evi-1 in the regulation of hematopoietic stem cells. J Cell Physiol., 222, 282-285.

Lee, J. Y., Huerta, P. T., Zhang, J., Kowal, C., Bertini, E., Volpe, B. T., & Diamond, B. (2009). Neurotoxic autoantibodies mediate congenital cortical impairments of offspring in maternal lupus. Nature Medicine, *15*, 91-96.

Li, Q., Lee, J.A. and Black, D.L. (2007). Neuronal regulation of alternative pre-mRNA splicing. Nat Rev Neurosci., 8, 819-831.

Llorian, M., Beullens, M., Andres, I., Ortiz, J.M. and Bollen, M. (2004). SIPP1, a novel pre-mRNA splicing factor and interactor of protein phosphatase-1. Biochem J., 378, 229-238.

Llorian, M., Beullens, M., Lesage, B., Nicolaescu, E., Beke, L., Landuyt, W., Ortiz, J.M. and Bollen, M. (2005). Nucleocytolasmic shuttling of the splicing factor SIPP1. J Bio Chem., 280, 38862-38869.

Mohler, H. (2007). Molecular regulation of cognitive functions and developmental plasticity: impact of $GABA_A$ receptors. J Neurochem., 102, 1-12.

Muto, Y., Krummel, D.P., Oubridge, C., Hernandez, H., Robinson, C.V., Neuhaus, D. and Nagai, K. (2004). The structure and biochemical properties of the human spliceosomal protein U1C. J Mol Bio., 341, 185-198.

Nelissen, R., Heinrichs, V., Habets, W.J., Siomons, F., Luhrmann, R. and van Venrooij, W.J. (1991). Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1snRNP. Nucleic Acids Res., 19, 449-454.

Patel, A.A. and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. Nat Rev Mol Cell Biol., 4, 960-970.

Rhoads, R.E. (1993). Regulation of eukaryotic protein synthesis by initiation factors. J Biol Chem., 268, 3017-3020.

Robinson, F.L., Niesman, I.R., Beiswenger, K.K. and Dixon, J.E. (2008). Loss of the inactive myotubularin-related phosphatase Mtmr13 leads to Charcot-Marie-Tooth 4B2-like peripheral neuropathy in mice. Proc Natl Acad Sci USA., 105, 4916-4921.

Sambrook, J. and Russell D.W. (2001). Molecular cloning: A laboratory manual. Third edition. Cold Spring Harbor Laboratory Press. New York.

Smoller, J.W., Rosenbaum, J.F., Biedman, J., Kennedy, J. and Dai, D., Racette, S.R., Larid, N.M., Kagan, J., Snidman, N., Hirshfeld-Becker, D., Tsuang, M.T., Sklar, P.B. and Slaugenhaupty, S.A. (2003). Association of genetic marker at the corticotropin-releasing hormone locus with behavioral inhibition. Biol Psychiatry., 54 1376-1381.

Song, W-J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., Loh, M., Felix, C., Roy, D-C., Busque, L., Kurnit, D., Willamn, C., Gerwitz, A.M., Speck, N.A., Bushweller, J.H., Li, F.P., Gardiner, K., Poncz, M., Maris, J.M. and Gilliland, D.G. (1996). Haploinsufficiency of CBF2 causes familial thrombocytopenia with propensity to develop acute meylogenous leukemia. Nat Genet., 23, 166-175.

Speck, NA. (2001). Core binding factor and its role in normal hematopoietic development. Curr Opin Hematol., 8, 192-196.

Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H.L., Ryner, L., Cheung, L.M., Erickson, C., Fisher, W.W., Greer, K., Hartouni, S., Howie, E., Jakkula, L., Joo, D., Wahl, M.C., Will, C.L. and Luhrmann, R. (2009). The spliceosome: design principle of dynamic RNP machine. Cell., 136, 701-718.

Walf, A.A. and Frye, C.A. (2007). The use of the elevated plus maze as an assay of anxiety-related behaviors in rodents. Nat Protoc., 2, 322-328.

Wiesner, S., Stier, G., Sattler, M. and Macias, M.J. (2002). Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. J Mol Biolo., 324, 807-822.

Wu, S., Ying. G., Wu, Q. and Capecchi M.R. (2007). Toward simpler and faster genome-wide mutagenesis in mice. Nat Genet., 39, 923-930.

van der Brug, M., Nalls, M.A. and Cookson, M.R. (2010) Deep sequencing of coding and non-coding RNA in CNS. Brain Res., 1338, 146-154.

CHAPTER 4

CONCLUSIONS

Chapter 2 dealt with the challenge of transferring a large chromosomal fragment between different species. It was attempted to develop the technology that enables us to move the chromosomal fragment from canine fibroblasts to mouse ES cells. As a model, the canine QTL that inversely regulates length and width of the femur growth was used. As stated in Chapter 2, there were two biological obstacles to overcome in order to achieve this objective. The first biological obstacle was to perform gene targeting in the canine somatic cells. The second biological obstacle was to pass the targeted canine chromosome with the QTL of interest through the cell membrane of the mouse ES cells. The first obstacle was exclusively dealt with in Chapter 2. The results presented in Chapter 2 showed that the targeting vector was indeed introduced into the canine immortalized fibroblasts, yet not targeted into the QTL correctly. Conventionally, gene targeting has been done in ES cells. However, availability of ES cells for each specie is limited. In those species without ES cells, gene targeting in the immortalized cells has been tried and improved. It has been shown that the immortalized cells can be targeted when the nuclear localization signaling sequence is incorporated into the targeting vectors (Bashir & Piedrahita, 2004; Meehan, et al., 2008). Additionally, it has been reported that homologous recombination occurs mainly in the late S/G2 phase and that targeting efficiency is enhanced in S-phase-synchronized cells by thymidine treatment

(Takata et al., 1998; Lundin et al., 2002). Thymidine can trap cells into the G_1/S-phase of cell cycles and the cell cycles can be synchronized in the majority of cells. When thymidine is removed from the cells, the cells are then allowed to proceed cell cycle through S-phase in a synchronized manner, which is when the synchronized cell should be transfected with the targeting vectors by electroporation. For a future project, the immortalized canine fibroblasts may be electroporated with the targeting vector containing the nuclear localization sequences combination of the cell cycle to see if the proper targeting can be achieved.

Gene targeting has been efficiently performed in mouse ES cells since this revolutionary genetic technology was invented. It has been reported that gene targeting has been successfully accomplished in ES cells and iPS cells from other species (Tong et al., 2010; Zou et al., 2009). Because canine ES cells carrying this specific QTL were unavailable, I attempted to make iPS cells from adult canine fibroblasts by using lentiviral and the *piggybac* transposon iPSC vectors. The results showed that partial iPS cells could be generated from the canine adult fibroblasts. However, those partial iPS cells were never reprogrammed into fully reprogrammed iPSCs. The major problem must be addressed to generate canine iPSCs and further succeed in gene targeting. One is the age of our canine fibroblasts. It has been demonstrated that the reprogramming is more difficult if the host cells are differentiated further along with the cell fates (Eminli et al., 2009). In this project the adult and neonatal canine fibroblasts were used for the canine iPSC formation. As the results showed, even the neonatal fibroblasts had difficulty to pass the initiation phase of iPSC generation. Thus, developmentally

immature canine cells (i.e., progenitors or embryonic cells of any cell types) with the QTL of interest should be reprogrammed to create the canine iPSCs.

The other major problem is compatibility of the mouse or human-derived reprogramming factors. In this study the mouse and human-derived reprogramming factors (*Oct4, Sox2, Klf4* and *c-Myc*) were used to generate the canine iPSCs. Although those four factors from either species were sufficient to generate iPS cells from any species except dogs, both mouse and human-derived reprogramming factors may not be compatible to activate the canine reprogramming machinery. In fact, a recent report demonstrated that canine iPS cell could be generated from the embryonic fibroblasts by using canine *Oct4, Sox2, Klf4* and *c-Myc* in combination with small molecule inhibitors (Shimada et al., 2010). Thus, the canine-derived reprogramming factors may be required to generate the canine iPSCs. Collectively, reprogramming the developmentally immature cells such as the embryonic fibroblasts by the canine-derived four stem cell factors may be tried to successfully generate the canine iPSCs and further achieve gene targeting in the canine cells.

In Chapter 3 the random mutagenesis of mouse genome by *piggybac* DNA transposon system provided the opportunity to characterize the novel gene, AK045681, expressed in mouse brain. The results showed that this gene is expressed specifically in neurons and the expression started only postnatally (at postnatal day 12). This gene is expressed in specific regions of the brain such as the limbic system (hippocampus and amydgala), neocortex, the cerebellum and olfactory system (olfactory bulb and anterior olfactory nucleus). Further analyses revealed that the gene is mainly expressed in GABAnergic neurons with an exception of granular cells in the cerebellum. Cellular

localization of the gene product was found to be nuclear and the gene was indeed translated in vitro. The timing of gene expression of this gene and no morphological deficit in the mutant brain suggested that the phenotypes might be found in behavior. Three behavioral tests were performed to identify phenotypes; 1) Metric and topological spatial learning task to check function of hippocampus; 2) Rota rod task to examine the cerebellar function; 3) Elevated plus maze to see normal anxiety behavior, which is regulated by the limbic system. The results of behavioral analyses showed that the mutant mice were less anxious or fearful than the wild type mice, which indicated that the gene might play a role in regulating the anxiety behavior in limbic system.

Based on the amino acid sequence similarity to spliceosomal protein U1C, the gene may be involved in RNA splicing. I hypothesized that the gene might be a component of the spliceosome or a splicing factor that is required to do alternative splicing for specific genes. Since the gene is expressed mainly in the GABAergic neurons, the gene might be involved in the alternative splicing of GABA receptors and transporters. Co-immunoprecipitation of the gene with the spliceosomal proteins (SmB/B'/N and U1-70K) was performed. Both spliceosomal proteins were key components of the spliceosome, and these proteins have shown to physically interact with U1C, whose amino acid sequence is similar to that of AK045681. However, Co-immunoprecipitation experiment revealed that AK did not physically interact with either spliceosomal proteins. Moreover, RT-PCR showed that all GABA receptors and transporter genes seemed to be spliced normally. Thus, the functions of AK045681 remain unknown. There are a few experiments to identify the functional roles of this gene as future directions of the project; 1) Immunopreiciptation of HA-tagged AK045681

to isolate binding partner(s) and mass spectrometry to identify the binding partner(s); 2) Co-immunoprecipitation of HA-tagged AK045681 with the identified binding partner to prove their physical association; 3) RNA-Seq to characterize the transcriptosome of neurons between the wild type littermate mice and the mutant mice, focusing on the information of splicing variants of known genes because biological function of AK045681 may be RNA splicing based on amino acid sequence similarity to U1C (Wang et al., 2009; Ozsolak & Milos, 2010; van der Brug et al., 2010). The further analyses of AK045681 mutant mice will provide additional information about the biology of anxiety-related behaviors and will potentially serve as the mouse model of the anxiety-related behaviors. Furthermore, the specific location and timing of AK045681 expression will be certainly utilized as a genetic tool including a postnatal neuronal specific Cre-driver.

References

Arbones, M.L., Austin, H.A., Capon, D.J. and Greenburg, G. (1994). Gene targeting in normal somatic cells: inactivation of the interferon-gamma receptor in myoblasts. Nature Genet., 6, 90-97.

Bashir, M. and Piedrahita, J.A. (2004). Nuclear localization signal and cell synchrony enhance gene targeting efficiency in primary fetal fibroblasts. Nucleic Acids Res., 32, e25.

Brown, J.P., Wei, W. and Sedivy, J.M. (1997). Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. Science, 277, 831-834.

Eminli, S., Foudi, A., Stadtfeld, M., Maherali, N., Ahfeldt, T., Mostoslavsky G., Hock, H. and Hochedlinger K. (2009). Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. Nat. Genet., 41, 968-978.

Hanson, K.D. and Sedivy, J.M. (1995). Analysis of biological selections for high-efficiency gene targeting. Mol. Cell. Biol., 15, 45-51.

Lundin, C., Erixon, K., Arnaudeau, C., Schultz, N., Jenssen, D., Meuth, M. and Helleday, T. (2002). Different roles for nonhomologous end joining and homologous recombination following replication arrest in mammalian cells. Mol. Cell. Biol., 22, 5869-5878.

Ozsolak, F. and Milos, F. (2010) RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet., 12, 87-98.

Shimada, H., Nakada, A., Hashimoto, Y., Shigeno, K., Shionoya, Y. and Nakamura, T. (2010) Generation of Canine induced pluripotent stem cells by retroviral transduction and chemical inhibitors. Mol. Reprod. Dev., 77, 1-2.

Takata, M., Sasaki, M.S., Sonda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. and Takeda, S. (1998). Homologous recombination and nonhomologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J., 17, 5497-5508.

Tong, C., Li, P., Wu, N.L., Yan, Y. and Ying, Q-L. (2010). Production of p53 gene knockout rats by homologous recombination in embryonic stem cells. Nature. 467, 211-213.

van der Brug, M., Nalls, M.A. and Cookson, M.R. (2010) Deep sequencing of coding and non-coding RNA in CNS. Brain Res., 1338, 146-154.

Waldman, A.S. (1992). Targeted homologous recombination in mammalian cells. Crit. Rev. Oncol. Hematol., 12, 49-64.

Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet., 10, 57-63.

Zou, J., Maeder, M.L., Mali, P., Pruett-Miller, S.M., Thibodeau-Beganny, S., Chou, B-K., Chen, G., Ye, Z., Park, I-H., Daley, G.Q., Porteus, M.H., Joung, J.K. and Cheng L. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell, 5, 97-110.